

1976 Census of Canada

Quality of Data

Series I: Sources of Error —
Sampling and Weighting

Recensement du Canada de 1976

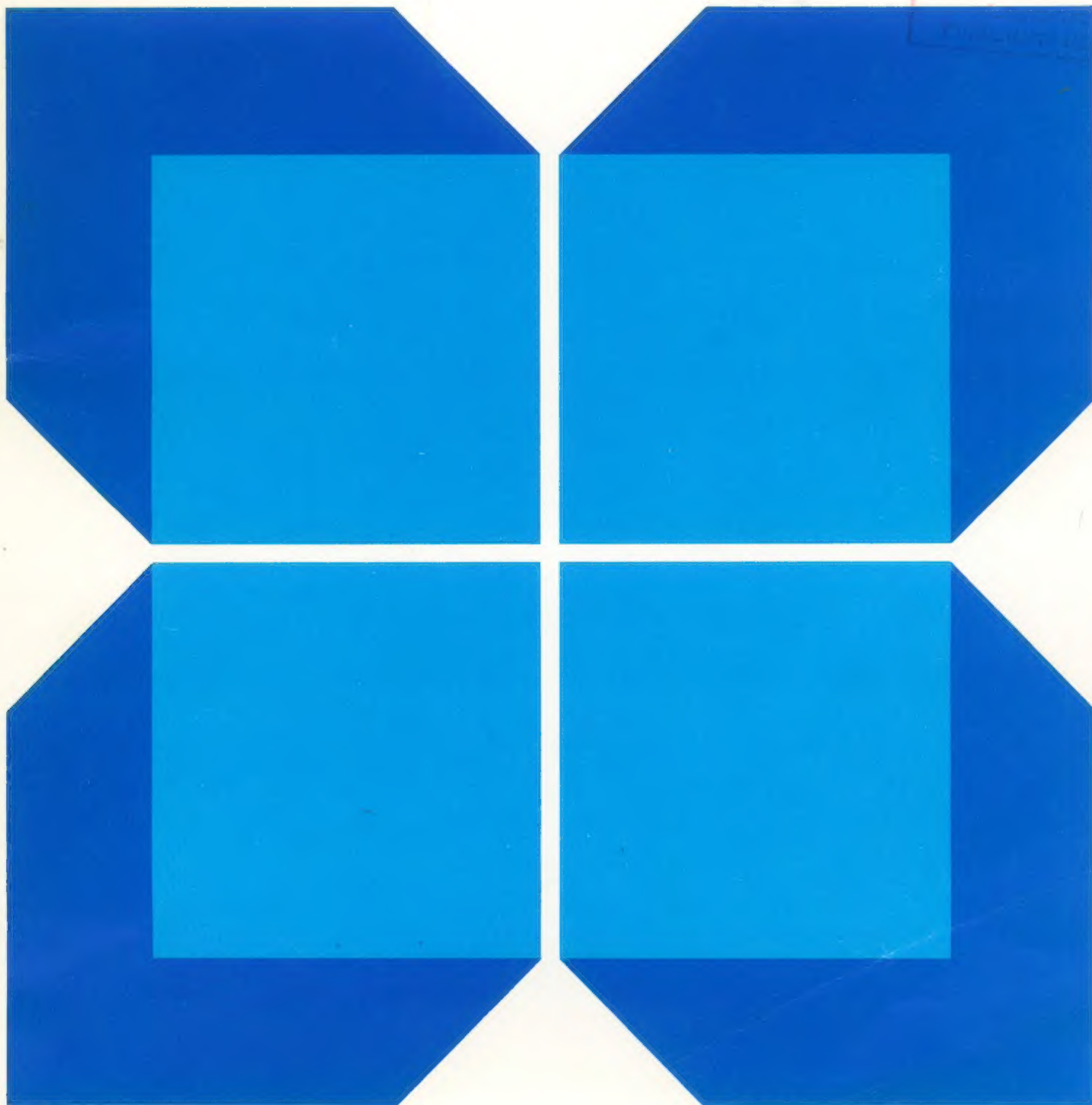
Qualité des données

Série I: Sources d'erreurs —
Échantillonnage et pondération

STATISTICS CANADA STATISTIQUE CANADA

M/P 5-1-

QUALITY OF DATA



Note

This and other government publications may be purchased from local authorized agents and other community bookstores or by mail order.

Mail orders should be sent to Publications Distribution, Statistics Canada, Ottawa, K1A 0T6, or to Publishing Centre, Supply and Services Canada, Ottawa, K1A 0S9.

Inquiries about this publication should be addressed to:

Census Survey Methods,
Census and Household Survey Methods Division,

Statistics Canada, Ottawa, K1A 0T6 (telephone: 995-8801) or to a local office of the bureau's User Advisory Services Division:

St. John's (Nfld.)	(726-0713)
Halifax	(426-5331)
Montréal	(283-5725)
Ottawa	(992-4734)
Toronto	(966-6586)
Winnipeg	(949-4020)
Regina	(569-5405)
Edmonton	(425-5052)
Vancouver	(666-3695)

Toll-free access to the regional statistical information service is provided in Nova Scotia, New Brunswick, and Prince Edward Island by telephoning 1-800-565-7192. Throughout Saskatchewan, the Regina office can be reached by dialing 1-800-667-3524, and throughout Alberta, the Edmonton office can be reached by dialing 1-800-222-6400.

Nota

On peut se procurer cette publication, ainsi que toutes les publications du gouvernement du Canada, auprès des agents autorisés locaux, dans les librairies ordinaires ou par la poste.

Les commandes par la poste devront parvenir à Distribution des publications, Statistique Canada, Ottawa, K1A 0T6, ou à Imprimerie et édition, Approvisionnements et services Canada, Ottawa, K1A 0S9.

Toutes demandes de renseignements sur la présente publication doivent être adressées à:

Méthodes d'enquête recensement,
Méthodes d'enquêtes ménages et recensement,

Statistique Canada, Ottawa, K1A 0T6 (téléphone: 995-8801) ou à un bureau local de la Division de l'assistance-utilisateurs situé aux endroits suivants:

St. John's (T.-N.)	(726-0713)
Halifax	(426-5331)
Montréal	(283-5725)
Ottawa	(992-4734)
Toronto	(966-6586)
Winnipeg	(949-4020)
Regina	(569-5405)
Edmonton	(425-5052)
Vancouver	(666-3695)

On peut obtenir une communication gratuite avec le service régional d'information statistique de la Nouvelle-Écosse, du Nouveau-Brunswick et de l'Île-du-Prince-Édouard en composant 1-800-565-7192. En Saskatchewan, on peut communiquer avec le bureau régional de Regina en composant 1-800-667-3524, et en Alberta, avec le bureau d'Edmonton au numéro 1-800-222-6400.

1976 Census of Canada

Recensement du Canada de 1976

Quality of Data

Series I: Sources of Error — Sampling and Weighting

Effect of Sampling and Weighting
Procedures on the Reliability of
1976 Census of Population and
Housing Data

Qualité des données

Série I: Sources d'erreurs — Echantillonnage et pondération

L'impact des procédures d'échantillonnage
et de pondération sur la fiabilité
des données du recensement de la
population et du logement de 1976

Published under the authority of
the President of the Treasury Board

Statistics Canada should be credited when
reproducing or quoting any part of this document

©Minister of Supply and Services, Canada, 1980

February 1980
8-1200-581

Price: \$3.00

Catalogue 99-844

Ottawa

Publication autorisée par
le président du Conseil du Trésor

Reproduction ou citation autorisée sous réserve
d'indication de la source: Statistique Canada

©Ministre des Approvisionnements et Services, Canada, 1980

Février 1980
8-1200-581

Prix: \$3.00

Catalògue 99-844

Ottawa

PREFACE

This bulletin covers the effects of sampling and weighting on the quality of census data. It explains the sampling and weighting procedures used in the 1976 Census, describes the studies that were undertaken to assess the effect of sampling and weighting on data quality, and presents an analysis of the results of these studies.

This bulletin was planned and produced as part of the 1976 Census Parametric Evaluation project in which the Census Characteristics Division and the Census Survey Methods Division were the main participants. The studies on which this bulletin was based were carried out by H. Arora (Census Survey Methods Division) and the text was prepared by G.J. Brackstone (Census Survey Methods Division).

PRÉFACE

Le présent bulletin illustre les effets de l'échantillonnage et de la pondération sur la qualité des données du recensement. Il explique les procédures d'échantillonnage et de pondération utilisées au recensement de 1976, décrit les études effectuées pour évaluer l'impact de l'échantillonnage et de la pondération sur la qualité des données et présente une analyse des résultats de ces études.

Le présent bulletin a été conçu et rédigé dans le cadre du projet d'évaluation paramétrique du recensement de 1976 dans lequel la Division des caractéristiques du recensement et la Division des méthodes d'enquête-recensement ont joué un rôle prédominant. Les études qui ont servi à la confection de ce bulletin ont été menées par H. Arora (Division des méthodes d'enquête-recensement) tandis que G.J. Brackstone (de la même division) s'est chargé de la rédaction.

TABLE OF CONTENTS

	Page
I. INTRODUCTION	9
II. SAMPLING IN CANADIAN CENSUSES ..	11
2.1 History of Sampling in the Canadian Census	11
2.2 The Decision to Use Sampling in the 1976 Census	12
2.3 The Sampling Scheme Used in the 1976 Census	13
2.4 Processing the Census Sample	14
III. ESTIMATION FROM THE CENSUS SAMPLE	15
3.1 Operational Considerations	15
3.2 Theoretical Considerations	16
3.3 Developing an Estimation Procedure for the Census Sample	17
3.4 The Raking Ratio Estimation Procedure	19
IV. THE SAMPLING AND WEIGHTING EVALUATION PROGRAM	23
V. EVALUATION OF SAMPLE SELECTION	25
5.1 Evaluation of the CR's Route	25
5.2 Comparison of Sample Households With All Households	27
5.3 Analysis and Summary	30
VI. EVALUATION OF WEIGHTING PROCEDURES AND SAMPLE POPULATION CONSISTENCY	33
6.1 Weighting Area Formation	33
6.2 Cross-classification and Collapsing ..	35
6.3 Sample Population Consistency	35
VII. THE CONTRIBUTION OF SAMPLING TO TOTAL VARIANCE	41
7.1 Total Variance	41
7.2 The Contribution of Sampling	42
7.3 Comparisons With Simple Random Sampling	43
VIII. CONCLUSION	47
Appendix	
I. Determining the Random Start	49
II. Weighting Matrices	55
III. Variables and Categories Used in the Statistical Tests	57

TABLE DES MATIÈRES

	Page
I. INTRODUCTION	9
II. APPLICATION DE L'ÉCHANTILLONNAGE LORS DE RECENSEMENTS CANADIENS	11
2.1 Historique de l'échantillonnage aux recensements canadiens	11
2.2 Décision d'appliquer la méthode d'échantillonnage au recensement de 1976	12
2.3 Plan de sondage utilisé lors du recensement de 1976	13
2.4 Dépouillement de l'échantillon du recensement	14
III. ESTIMATION BASÉE SUR L'ÉCHANTILLON DU RECENSEMENT	15
3.1 Considérations d'ordre pratique	15
3.2 Considérations d'ordre théorique	16
3.3 Élaboration de méthodes d'estimation de l'échantillon du recensement	17
3.4 Méthode itérative d'estimation des quotients	19
IV. PROGRAMME D'ÉVALUATION DE L'ÉCHANTILLONNAGE ET DE LA PONDÉRATION	23
V. ÉVALUATION DE LA MÉTHODE DE SÉLECTION DE L'ÉCHANTILLON	25
5.1 Évaluation de l'itinéraire du RR	25
5.2 Comparaison des ménages de l'échantillon avec l'ensemble des ménages ...	27
5.3 Analyse et résumé	30
VI. ÉVALUATION DES PROCÉDURES DE PONDÉRATION ET DE LA COMPATIBILITÉ DE L'ÉCHANTILLON ET DE LA POPULATION	33
6.1 Formation des régions de pondération	33
6.2 Classement recoupé et agrégation	35
6.3 Compatibilité de l'échantillon et de la population	35
VII. INCIDENCE DE L'ÉCHANTILLON SUR LA VARIANCE TOTALE	41
7.1 Variance totale	41
7.2 Incidence de l'échantillonnage	42
7.3 Comparaison avec l'échantillon aléatoire simple	43
VIII. CONCLUSION	47
Appendice	
I. Détermination aléatoire des points de départ	49
II. Matrices de pondération	55
III. Variables et catégories utilisées lors des tests statistiques	57

TABLE OF CONTENTS – Concluded

Appendix	Page
IV. Differences Between Census Totals and Corresponding Sample Estimates for Selected Categories, by Province	59
References	69

TABLE DES MATIÈRES – fin

Appendice	Page
IV. Écarts entre les totaux du recensement et les estimations échantillonnales correspondantes de catégories choisies, par province	59
Références	69

LIST OF TABLES

	Page
1. Number of EAs Selected in Each Stratum for the Evaluation of Sample Selection Procedures	26
2. Means and Proportions Among (a) Private Households in the Sample and (b) All Private Households, for a Random Sample of 1,000 EAs	29
3. Number (and Percentage) of EAs Having at Least One Significant Test Result at the 5% Level by Variable and by Route Classification (Based on 696 Urban EAs) ..	30
4. Formation of WAs from CTs in Selected Municipalities Within CMAs/CAs	34
5. Most Frequent Collapsing Patterns	36
6. Differences Between Census Totals and Corresponding Sample Estimates for Selected Categories, for Canada	37
7. CDs by Size of Sample Population Difference for Various Categories for the 46 CDs in the Atlantic Provinces	39
8. Percentage Distribution of EAs by Size of Sample Population Difference for Various Categories for EAs in the Atlantic Provinces	39
9. Estimated Breakdowns of Total Variance into (a) Sampling and Simple Response Variance and (b) Correlated Response Variance, for Selected Categories	43
10. Comparison of Estimated Sampling Error for a Selection of Census Categories With the Sampling Error Assuming Simple Random Sampling	45

LISTE DES TABLEAUX

	Page
1. Nombre de SD choisis dans chaque strate pour l'évaluation des méthodes de sélection des échantillons	26
2. Moyennes et proportions a) des ménages et des logements privés dans l'échantillon par rapport à b) l'ensemble des ménages et des logements privés, dans un échantillon aléatoire de 1,000 SD	29
3. Nombre (et pourcentage) de SD pour lesquels on a obtenu au moins un résultat significatif du test à 5 % selon certaines variables et selon la catégorie d'itinéraire (Basé sur 696 SD urbains)	30
4. Formation des RP à partir de SR dans les municipalités choisies à l'intérieur des RMR/AR. . .	34
5. Patterns de regroupement les plus fréquents . . .	36
6. Écarts entre les totaux du recensement et les estimations échantillonnales correspondantes de catégories choisies au niveau du Canada	37
7. DR selon l'écart entre l'échantillon et la population pour diverses catégories des 46 DR des provinces de l'Atlantique	39
8. Répartition en pourcentage des SD selon l'écart entre l'échantillon et la population pour diverses catégories des SD des provinces de l'Atlantique	39
9. Estimation de la répartition de la variance totale en fonction a) de la variance échantillonnale et de la variance de réponse simple et b) de la variance de réponse corrélée, pour certaines catégories	43
10. Comparaison de l'erreur échantillonnale estimée pour certaines catégories du recensement avec l'erreur échantillonnale dans l'hypothèse d'un échantillonnage aléatoire simple	45

I. INTRODUCTION

Sampling is an accepted practice in many aspects of life today. The quality of produce in a market may be judged visually by a sample before a purchase is made; we form opinions about people based on samples of their behaviour; we form impressions about countries or cities based on brief visits to them. These are all examples of sampling in the sense of drawing inferences about the "whole" from information for a "part".

In a more scientific sense sampling is used, for example, by accountants in auditing financial statements, in industry for controlling the quality of items coming off a production line, and by the takers of opinion polls and surveys in producing information about a population's views or characteristics. In general, the motivation to use sampling stems from a desire either to reduce costs or to obtain results faster, or both. The disadvantage of sampling is that the results based on a sample may not be as precise as those based on the whole population. However, when the loss in precision (which may be quite small when the sample is large) is tolerable in terms of the uses to which the results are to be put, the use of sampling may be cost-effective. Furthermore, the reduction in the scale of a study achieved through using sampling may in fact lead to a reduction in errors from non-sampling sources, thus compensating to some extent for the loss of precision resulting from sampling.

The 1976 Census of Population and Housing made use of sampling in a variety of ways. It was used in ensuring that the quality of the Census Representative's work in collecting questionnaires met certain standards; it was used in the control of the quality of coding responses during office processing; it was used in determining how many persons away from home on census night were not enumerated at their usual residence and therefore needed to be added into the census population; it was used in evaluating the quality of census data. However, the primary use of sampling in the census was during the field enumeration when all but the basic census data were collected only from a sample of households. This bulletin describes this last use of sampling and evaluates the effect of sampling on the quality of census data.

I. INTRODUCTION

De nos jours, l'échantillonnage constitue une technique largement utilisée dans maints domaines. Nous pouvons juger de la qualité d'un produit à l'aide d'un échantillon avant d'en faire l'achat; nous nous formons une opinion des gens sur des échantillons de leur comportement; nos impressions sur tel pays ou telle ville sont basées sur les courts voyages que nous y avons faits. Ce sont là des exemples d'échantillonnage menant à tirer des conclusions sur le "tout" à partir de connaissances portant sur une "partie".

Scientifiquement parlant, l'échantillonnage est utilisé par les comptables, par exemple, lorsqu'ils effectuent la vérification des états financiers; dans l'industrie, pour vérifier la qualité d'articles produits en série et par les initiateurs de sondage d'opinion et d'enquêtes pour recueillir des renseignements sur la façon de penser ou les caractéristiques d'une population donnée. On a généralement recours à l'échantillonnage pour la réduction de coûts ou pour obtenir rapidement des données, ou les deux. Le désavantage de cette méthode est que les résultats obtenus à partir d'un échantillon peuvent ne pas être aussi précis que ceux recueillis auprès de la population totale. Toutefois, l'application d'un sondage peut être avantageuse quand la perte de précision (qui peut être minime si l'échantillon est suffisamment grand) est acceptable en fonction de l'utilisation prévue des résultats. De plus, la réduction de l'envergure d'une étude faite à partir d'un échantillon peut mener à la réduction des erreurs non dues à l'échantillonnage, ce qui compense en partie le manque de précision attribuable à l'échantillonnage.

Pour le recensement de la population et du logement de 1976, on a appliqué les méthodes d'échantillonnage de plusieurs façons. On y a eu recours pour s'assurer que la qualité du travail des représentants du recensement dans le cadre de la collecte des questionnaires respectait certaines normes; on y a eu recours pour le contrôle qualitatif du codage des réponses au cours du dépouillement dans les bureaux régionaux; on y a eu recours pour déterminer combien de personnes, absentes de leur domicile la nuit du recensement, n'ont pas été dénombrées à leur résidence habituelle et ont, par conséquent, dû être ajoutées à la population du recensement; on y a eu recours pour évaluer la qualité des données du recensement. Toutefois, la principale application de l'échantillonnage au recensement a été effectuée pendant le dénombrement sur le terrain alors que toutes les données, excepté les données fondamentales, ont été recueillies auprès d'un échantillon de ménages. Le présent bulletin décrit cette dernière utilisation de l'échantillonnage et présente une évaluation de l'incidence de cette méthode sur la qualité des données du recensement.

Chapter II reviews the history of the use of sampling in Canadian censuses and describes the sampling procedures used in the 1976 Census, while Chapter III explains the procedures used for weighting up the sample data to the population level. In Chapter IV, the program of studies designed to evaluate the 1976 Census sampling and weighting procedures is presented, while Chapters V, VI and VII present the results of different aspects of this evaluation.

Le chapitre II donne l'historique de l'application de l'échantillonnage aux recensements du Canada et la description des méthodes d'échantillonnage utilisées lors du recensement de 1976. Le chapitre III expose les méthodes utilisées pour la pondération des données échantillonnales en fonction des chiffres totaux de la population. Le chapitre IV expose le programme d'évaluation de l'échantillonnage et de la pondération du recensement de 1976, et les chapitres V, VI et VII donnent un exposé des résultats de divers aspects de cette évaluation.

II. SAMPLING IN CANADIAN CENSUSES

In the context of a Census of Population and Housing, sampling refers to the process whereby certain characteristics are collected and processed only for a random sample of the dwellings and persons identified in the complete census enumeration. Tabulations that depend on characteristics collected only on a sample basis are then obtained for the whole population by scaling up the results for the sample to the full population level. Characteristics collected on all dwellings or persons in the census will be referred to as "basic characteristics" while those collected only on a sample basis will be known as "sample characteristics".

2.1 History of Sampling in the Canadian Census¹

Sampling was first used in the Canadian census in 1941. A Housing Schedule was completed for every tenth dwelling in each census subdistrict. The information from 27 questions on the separate Housing Schedule was integrated with the data in the personal and household section of the Population Schedule for the same dwelling, thus allowing cross-tabulation of sample and basic characteristics. Also in the 1941 Census, sampling was used at the processing stage to obtain early estimates of earnings of wage-earners, of the distribution of the population of working ages, and of the composition of families in Canada. In this case, a sample of every tenth enumeration area across Canada was selected and all Population Schedules in these areas were processed in advance.

Again in 1951, the Census of Housing was conducted on a sample basis. This time every fifth dwelling (those whose identification numbers ended in a 2 or 7) was selected to complete a housing document containing 24 questions. In the 1961 Census, persons 15 years of age and over in a 20% sample of private households were required to complete a Population Sample Questionnaire containing questions on internal migration, fertility and income. Sampling was not used in the smaller censuses of 1956 and 1966.

The 1971 Census saw several major innovations in the method of census-taking. The primary change was from the traditional canvasser method of enumeration to the use of self-enumeration for the majority of the population. This change was prompted by the

¹ More detailed information for specific censuses can be found in the Administrative Report of the appropriate census. References to these reports can be found at the end of this bulletin.

II. APPLICATION DE L'ÉCHANTILLONNAGE LORS DE RECENSEMENTS CANADIENS

Dans un recensement de la population et du logement, l'échantillonnage consiste à recueillir et à traiter certaines caractéristiques à partir d'un échantillon aléatoire de logements et de personnes dénombrés au cours du recensement. Les totalisations pour ces caractéristiques pour l'ensemble de la population sont ensuite obtenues en pondérant les données échantillonnales pour ces caractéristiques en fonction de coefficients établis par le recensement de caractéristiques fondamentales.

2.1 Historique de l'échantillonnage aux recensements canadiens¹

L'échantillonnage a été utilisé pour la première fois au cours du recensement de 1941. Un questionnaire sur le logement a été remis à tous les 10 logements de chaque sous-district de recensement. Les réponses aux 27 questions de ce questionnaire ont été intégrées à celles du questionnaire sur la population et les ménages correspondant à ces logements, ce qui a permis de faire le recoupement des caractéristiques de l'échantillon et des caractéristiques de base. Lors de ce même recensement, on a utilisé l'échantillonnage au cours de l'étape du dépouillement pour obtenir des estimations provisoires du revenu des salariés, de la répartition de la population active selon l'âge et de la composition des familles au Canada. Dans ce cas, l'échantillon était constitué du dixième de tous les secteurs de dénombrement du Canada et tous les questionnaires sur la population de ces secteurs ont été dépouillés à l'avance.

L'échantillonnage a également été utilisé pour le recensement du logement de 1951. Cette fois, on a choisi chaque cinquième logement (ceux dont le numéro d'identification se terminait par 2 ou par 7) et le questionnaire sur le logement comportait 24 questions. Au cours du recensement de 1961, on a prélevé un échantillon de 20 % des ménages privés dont les membres âgés de 15 ans et plus ont été priés de remplir un questionnaire — échantillon sur la population comportant des questions sur la migration interne, la fécondité et le revenu. La méthode d'échantillonnage n'a pas été utilisée au cours des recensements de moindre envergure de 1956 et 1966.

Plusieurs innovations importantes ont été apportées aux méthodes de recensement en 1971. La première a été l'utilisation de l'autodénombrement pour la majorité de la population au lieu de la méthode traditionnelle de recensement par représentant. Ce changement repose sur

¹ Des renseignements détaillés sur les recensements particuliers sont contenus dans le rapport administratif des recensements en question. Ces rapports sont mentionnés à la fin de ce bulletin.

results of several studies in Canada and elsewhere (Fellegi, 1964; Hansen et al., 1959) that indicated that the effect of the enumerator was a major contribution to the variance² of census figures in a canvasser census. Thus the use of self-enumeration was expected to reduce the variance of census figures through reducing the effect of the enumerator, while at the same time giving the respondent more time and privacy in which to answer the census questions – factors which might also be expected to yield more accurate responses.

The second aspect of the 1971 Census that differentiated it from any earlier census was its content. The number of topics covered and the number of questions asked were greater than in any previous Canadian census. Considerations of cost, respondent burden, and timeliness versus the level of data quality to be expected using self-enumeration and sampling led to a decision to collect all but certain basic characteristics on a one-third sample basis in the 1971 Census. In all but the more remote areas of Canada, every third private household received the "long form" which contained all the census questions while the remaining private households received the "short form" containing only the basic questions covering name, relationship to head, sex, date of birth, marital status, mother tongue, type of dwelling, tenure, number of rooms, water supply, toilet facilities, and certain coverage items. All households in pre-identified remote enumeration areas and all collective dwellings³ received the long form.

A more detailed description of the consideration of the use of sampling in the 1971 Census is given in *Sampling in the Census* (Dominion Bureau of Statistics, August 1968). For a description of the use of sampling in censuses in the United States, see U.S. Bureau of the Census, Technical Paper No. 13, 1965.

2.2 The Decision to Use Sampling in the 1976 Census

The content of the 1976 Census was considerably less than that of the 1971 Census. Furthermore, the 1976 Census did not include the questions that cause the most difficulty in collection (e.g., income) or that are costly to code (e.g., occupation, industry and place of work). Therefore, the benefits of sampling in terms of cost savings and reduced respondent burden

les résultats de plusieurs études effectuées au Canada et ailleurs (Fellegi, 1964, Hansen et associés, 1959) dont on a pu déduire que la présence d'un représentant avait un effet important sur la variance² des chiffres du recensement. On prévoyait que l'autodénombrement permettrait de réduire la variance des chiffres du recensement attribuable à la présence du représentant et, en même temps, donnerait au répondant le temps voulu pour remplir le questionnaire à son aise; par voie de conséquence, on pouvait s'attendre que les réponses seraient plus exactes que par le passé.

Un autre changement survenu dans le recensement de 1971 a été la modification du contenu. Le nombre de sujets traités et de questions posées était plus élevé qu'aux recensements précédents. Une étude des coûts, du fardeau du répondant et des délais de production par rapport à la qualité des données dans le cadre d'un recensement par autodénombrement a présidé à la décision de recueillir toutes les données, sauf celles portant sur quelques caractéristiques de base (appelées caractéristiques à 100 % ou recensées), auprès d'un échantillon d'un tiers de la population au recensement de 1971 (appelées caractéristiques à 33 % ou échantillonnelles). Sauf dans les régions les plus isolées du Canada, le tiers des ménages a reçu le questionnaire complet comportant toutes les questions du recensement, tandis que le reste des ménages a reçu le questionnaire abrégé qui ne contenait que les questions de base sur le nom, le lien de parenté avec le chef de ménage, le sexe, la date de naissance, l'état matrimonial, la langue maternelle, le type de logement, le mode d'occupation, le nombre de pièces, l'alimentation en eau, les toilettes ainsi que certaines questions en rapport avec la couverture du recensement. Tous les ménages vivant dans des régions prédésignées et dans des logements collectifs³ ont reçu le questionnaire complet.

Le rapport statistique "Échantillonnage en tant que technique de recensement" publié en août 1968 donne une description détaillée des projets d'application de cette méthode au recensement de 1971. Pour la description détaillée de l'utilisation de l'échantillonnage dans les recensements aux États-Unis, voir le rapport technique n° 13, 1965, du *United States Bureau of the Census*.

2.2 Décision d'appliquer la méthode d'échantillonnage au recensement de 1976

Le contenu du recensement de 1976 était beaucoup moindre que celui de 1971. En outre, le questionnaire de 1976 ne contenait pas les questions qui posent le plus de difficulté à recueillir, par exemple, les questions sur le revenu, ni celles dont les frais de codage sont les plus élevés, par exemple, celles sur la profession, l'activité économique et le lieu de travail. Par conséquent, les avantages de

² The "variance" of an estimate is a measure of its precision. Variance is discussed more fully in Chapter VII.

³ A collective dwelling is a dwelling in which 10 or more persons unrelated to the head of the household are likely to reside. Examples include hotels, hospitals, institutions and work camps.

² La "variance" d'une estimation est une mesure de sa précision. La variance est expliquée plus en détail au chapitre VII.

³ Un logement collectif est un logement où habitent 10 personnes ou plus qui n'ont aucun lien de parenté avec le chef de ménage, par exemple, les hôtels, les hôpitaux, les institutions et les baraquements.

were less clear than for the 1971 Census. Nevertheless, after estimating the potential cost savings to be expected with various sampling fractions, and considering the public relations issues related to a reversion to 100% enumeration after a successful application of sampling in 1971, it was decided to use the same sampling procedure in 1976 as in 1971.

2.3 The Sampling Scheme Used in the 1976 Census

Data for more than 98% of the population were collected by self-enumeration in the 1976 Census. In all self-enumeration areas, a 1 in 3 systematic sample of occupied private dwellings was selected to receive a long form (Form 2B) containing all census questions. The remaining private dwellings received a short form (Form 2A) containing only the basic census questions (name, relationship to head, sex, marital status, mother tongue, date of birth, type of dwelling, tenure, and coverage questions). Within each enumeration area (EA), a random start (1, 2 or 3) was specified to indicate which was the first dwelling to receive the long form.

All dwellings in those areas enumerated by the canvasser method received the Form 2B. All collective dwellings also received the Form 2B. Information on unoccupied private dwellings was recorded on a Form 2A.

In sampling terminology, the sample can be described as a stratified systematic sample of private occupied dwellings using a constant 1 in 3 sampling rate in all strata (enumeration areas). As a sample of persons, it can be regarded as a stratified systematic cluster sample with dwellings as clusters. (For a more detailed description of the concepts and terminology of sampling, see Stuart, 1976, or Cochran, 1977.)

The basic drop-off or delivery procedure required the Census Representative (CR) to pre-plan a route covering all dwellings in his/her EA following certain rules about the order in which city blocks and rural roads were to be enumerated, and then to visit each dwelling and leave a census questionnaire. The selection of the sample, i.e., the decision as to which type of questionnaire to leave at each occupied dwelling, was facilitated by the Visitation Record (VR), the document in which the CR listed each dwelling in his/her area. This document was printed so that every third line was shaded to signify that a Form 2B should be delivered. The random start was implemented by deleting either zero, one or two lines at the start of the VR according to whether the third, second or first

l'échantillonnage par rapport aux coûts et à l'allègement du fardeau du répondant étaient moins évidents qu'au recensement de 1971. Néanmoins, après avoir fait l'estimation des économies éventuelles reliées à diverses fractions de sondage et étudié les conséquences sur le plan des relations publiques du retour au dénombrement intégral, il a été décidé d'appliquer en 1976 les mêmes méthodes d'échantillonnage qu'en 1971.

2.3 Plan de sondage utilisé lors du recensement de 1976

Au recensement de 1976, on a recueilli des données sur plus de 98 % de la population par autodénombrement. Un échantillon systématique du tiers des logements privés occupés a été choisi dans tous les secteurs d'autodénombrement pour remplir le questionnaire complet (formule 2B) contenant toutes les questions du recensement. Les autres logements privés ont reçu le questionnaire abrégé (formule 2A) qui ne contenait que les questions fondamentales de recensement, comme le nom, le lien avec le chef de ménage, le sexe, l'état matrimonial, la langue maternelle, la date de naissance, le type de logement, le mode d'occupation ainsi que les questions en rapport avec la couverture du recensement. Dans chaque secteur de dénombrement (SD), le numéro du premier logement (1, 2 ou 3) devant recevoir le questionnaire complet a été choisi au hasard.

Tous les logements des secteurs dénombrés par un représentant du recensement ainsi que tous les logements collectifs ont reçu la formule 2B. Les données sur les logements privés vacants ont été inscrites sur une formule 2A.

Dans la terminologie de l'échantillonnage, l'échantillon est défini comme un échantillon stratifié systématique de logements privés occupés, choisi en fonction d'un taux d'échantillonnage constant de 1 sur 3 dans toutes les strates (secteurs de dénombrement). Dans le cas d'un échantillon de personnes, il peut être défini comme étant un échantillon stratifié systématique en grappes, un logement constituant une grappe. Stuart (1976) et Cochran (1977) donnent une description détaillée des concepts et de la terminologie se rapportant à l'échantillonnage.

D'après les procédures de livraison, le représentant du recensement (RR) devait planifier un itinéraire couvrant tous les logements de son SD, en se conformant aux règles établies pour l'ordre dans lequel les îlots urbains et les routes rurales devaient être dénombrés; il devait ensuite se rendre à chaque logement et y laisser un questionnaire. Le choix de l'échantillon, c'est-à-dire la décision sur le type de questionnaire à laisser dans chaque logement occupé, a été fait à partir du Registre des visites (RV) dans lequel le RR inscrivait tous les logements de son secteur. Chaque troisième ligne du registre était ombrée, indiquant qu'une formule 2B devait être livrée. Le point de départ aléatoire était établi à partir de la première ligne ou en sautant une ou deux lignes, ce qui déterminait l'ordre dans lequel le questionnaire complet

dwelling in the EA was to be the first to receive the long form (see Appendix I). Thereafter, the dwelling listed on each shaded line automatically received the long form. These procedures were spelled out in the CR's Manual and emphasized in his/her training in order to minimize the risk of any deviation from the specified procedure for selecting the sample.

2.4 Processing the Census Sample

Once the CR had obtained the completed questionnaire (Form 2A or 2B) from each dwelling in his/her area and this work had been approved, the questionnaires were sent for manual processing in one of eight Regional Offices and then to Ottawa for computer processing. All the Forms 2A and 2B for one EA were kept together until all the data for the EA had been successfully converted to magnetic tape. At this point, all information pertaining to basic characteristics, from both Forms 2A and Forms 2B, was extracted from magnetic tape and passed on to the edit and imputation system for processing through to a final census 100% data base from which published census figures for basic characteristics were derived. The weighting operation, described in detail in Chapter III, was also performed using the 100% data base, although the weights themselves were used only for retrieving data from the sample data base.

Once the 100% data had been finalized, the raw sample data were extracted from the magnetic tapes and linked with the now edited 100% data for the same dwelling. This linked file therefore contained, for each dwelling enumerated on a Form 2B, the basic data in their final imputed form⁴ and the sample data in their raw form as captured directly from the questionnaires. This file became the input to the edit and imputation system for sample data which led to a final census sample data base. This data base also contained the household weight for each household in the sample, and the person weight for each person in the sample calculated using the 100% data base as described in Chapter III.

⁴ It also contained some raw basic data for evaluation purposes.

devait être livré, c'est-à-dire à partir du troisième, deuxième ou premier logement (voir l'appendice I). Par la suite, chaque logement inscrit sur une ligne ombrée recevait automatiquement le questionnaire complet. Ces procédures étaient expliquées dans le Manuel du RR et une attention particulière y était portée pendant la formation du RR afin de minimiser tout risque d'erreur dans la sélection de l'échantillon.

2.4 Dépouillement de l'échantillon du recensement

Une fois que le RR avait reçu tous les questionnaires remplis (formule 2A ou 2B) de tous les logements de son SD et que son travail avait été approuvé, ces formules étaient envoyées à l'un des huit bureaux régionaux pour le dépouillement manuel, puis à Ottawa pour le traitement mécanographique. Toutes les formules 2A et 2B d'un même SD étaient conservées ensemble jusqu'au moment où toutes les données sur le SD étaient enregistrées sur bande magnétique. À ce stade, toutes les données des deux formules portant sur les caractéristiques de base étaient extraites de la bande magnétique et transférées au système de contrôle et d'imputation pour fins de traitement et de production d'une base de données à 100 %; les données sur les caractéristiques à 100 % figurant dans les publications du recensement ont été tirées de cette base. La procédure de pondération, décrite en détail au chapitre III, a également été appliquée en utilisant la base des données à 100 % bien que les poids eux-mêmes ne furent utilisés que pour les données à 33 %.

Une fois le traitement des données à 100 % terminé, les données brutes de l'échantillon étaient extraites des bandes magnétiques et appariées aux données à 100 % maintenant vérifiées se rapportant aux mêmes logements. Ainsi, le nouveau fichier contenait, pour chaque logement dénombré sur une formule 2B, les données à 100 % déjà imputées⁴ et les données brutes de l'échantillon telles qu'elles avaient été extraites des questionnaires. Ce fichier a servi de base au système de contrôle et d'imputation des données échantillonnelles qui a abouti à la constitution du fichier définitif des données échantillonnelles du recensement. Ce fichier contenait également le coefficient de pondération de chaque ménage et celui de chaque personne dans l'échantillon, calculés selon la méthode décrite au chapitre III.

⁴ Ce fichier contenait également quelques données à 100 % brutes pour fins d'évaluation.

III. ESTIMATION FROM THE CENSUS SAMPLE

Any sampling procedure requires an associated estimation procedure for scaling sample data up to the full population level. The choice of an estimation procedure is generally governed by both operational and theoretical constraints. From the operational viewpoint, the procedure must be feasible within the processing system of which it is a part, while from the theoretical viewpoint the procedure should minimize the sampling error of the estimates it produces. In the following two sections, the operational and theoretical considerations relevant to the choice of estimation procedures for the census sample are described.

3.1 Operational Considerations

Mathematically, an estimation procedure can be described by an algebraic formula that shows how the value of the estimator for the population is calculated as a function of the observed sample values.

In small surveys that collect only one or two characteristics, or in cases where the estimation formula is very simple, it might be possible to calculate the sample estimates by applying the given formula to the sample data for each estimate required. However, in a survey or census in which a wide range of characteristics is collected, or in which the estimation formula is at all complex, the procedure of applying a formula separately for each estimate required is not feasible. In the case of a census for example, every cell of every tabulation based on sample data at every geographic level represents a sample estimate which under this approach would require a separate application of the estimation formula. In addition, the calculation of each estimate separately would not necessarily lead to consistency between the various estimates made from the same census sample.

Therefore, the approach taken in the census (and in many sample surveys) is to split the estimation procedure into two stages: (a) the calculation of weights (known as the weighting procedure); (b) the summing of weights to produce estimated population counts.

Any mathematical complexity is then contained in step (a) which is performed just once, while step (b) is reduced to a simple process of summing weights which takes place at the time a tabulation is retrieved. Also, since the weight attached to each sample unit is the same for whatever tabulation is being retrieved, consistency between different estimates based on sample data is assured.

III. ESTIMATIONS BASÉES SUR L'ÉCHANTILLON DU RECENSEMENT

Tout échantillonnage nécessite l'application de techniques d'estimation pour pondérer les données échantillonnales en vue d'obtenir des estimations à l'échelle de la population totale. Les techniques d'estimation sont généralement choisies en fonction d'exigences d'ordre pratique et théorique. Sur le plan pratique, cette méthode doit être compatible avec le système de dépouillement auquel elle s'intègre, tandis que sur le plan théorique, elle doit minimiser l'erreur due à l'échantillonnage des estimations qu'elle produit. Les deux chapitres qui suivent décrivent les considérations pratiques et théoriques qui régissent le choix des méthodes d'estimation de l'échantillon du recensement.

3.1 Considérations d'ordre pratique

Une procédure d'estimation peut être décrite mathématiquement à l'aide d'une formule algébrique qui indique comment la valeur de l'estimateur de la population est calculée en fonction des valeurs observées dans l'échantillon.

Quand il s'agit de petites enquêtes qui ne portent que sur une ou deux caractéristiques, ou, lorsque la formule d'estimation est très simple, il est possible de calculer les estimations échantillonnales en appliquant la dite formule aux données échantillonnales pour chaque estimation requise. Toutefois, dans le cas d'une enquête ou d'un recensement portant sur une gamme étendue de caractéristiques, ou lorsque la formule d'estimation est assez complexe, il n'est pas possible d'appliquer une formule séparément pour chaque estimé désiré. Dans le cas du recensement, par exemple, chaque cellule de chaque totalisation est compilée sur des données échantillonnales recueillies à chaque niveau géographique et représente une estimation échantillonnale qui nécessite une application distincte de la formule d'estimation. Par ailleurs, le fait de calculer séparément chaque estimation ne mènerait pas nécessairement à la compatibilité des diverses estimations effectuées à partir d'un même échantillon.

Par voie de conséquence, la méthode utilisée pour le recensement et pour de nombreuses enquêtes par sondage consiste à décomposer le processus d'estimation en deux étapes: a) le calcul du coefficient de pondération, c'est-à-dire la pondération elle-même et b) la sommation des coefficients de pondération pour produire des chiffres estimatifs de la population.

Les seules difficultés mathématiques, s'il en est, se rapportent à l'étape a) qui n'est effectuée qu'une fois, tandis que l'étape b) se limite à un simple processus de sommation des coefficients de pondération au moment où une totalisation est extraite. De plus, comme le coefficient de pondération relié à chaque unité échantillonnale est le même, quel que soit la totalisation effectuée, la compatibilité entre les diverses estimations basées sur les données de l'échantillon est assurée.

3.2 Theoretical Considerations

For a given sample design and a given estimation procedure, one can, from sampling theory, make a statement about the chances that a certain interval about the estimate will contain the unknown population value being estimated. The primary criterion in the choice of an estimation procedure is minimization of the width of such intervals so that these statements about the unknown population values are as precise as possible. The usual measure of precision for comparing estimation procedures is known as standard error. As well as minimizing standard error, a second objective in the choice of estimation procedure for the census sample is to ensure, as far as possible, that sample estimates for basic characteristics are consistent with the corresponding known population values. Fortunately, these two objectives are usually complementary in the sense that sampling error tends to be reduced by ensuring that sample estimates for certain basic characteristics are consistent with the corresponding population figures. However, forcing sample estimates for basic characteristics to be consistent with corresponding population figures for very small cells could have a detrimental effect on the sampling error of estimates for the sample characteristics themselves.

In the absence of any information about the population being sampled other than that collected for sample units, the estimation procedure would be restricted to weighting the sample units inversely to their probabilities of selection (e.g., if all units in one province had a one in 10 chance of selection, then all selected units in that province would receive a weight of 10). However, in practice, one almost always has some supplementary knowledge about the population (e.g., its total size, and possibly its breakdown by a certain variable — perhaps by province). Such information can be used to improve the estimation formula so as to produce estimates with a greater chance of lying close to the unknown population value. In the case of the census sample, a large amount of very detailed information about the population being sampled is available in the form of the basic 100% data at every geographic level. On the one hand, we can take advantage of this population information to improve the estimates made from the census sample; on the other hand, this wealth of information can also be an embarrassment in the sense that it is impossible to make the sample estimates consistent with all the population information at every geographic level. Differences between sample estimates and population values become visible when a cross-tabulation of a sample variable and a basic variable is produced. The tabulation has to be based on sample data with the result that the marginal totals for the basic variable are sample estimates that can be compared with the corresponding population figures appearing in a dif-

3.2 Considérations d'ordre théorique

Selon la théorie de l'échantillonnage, il est possible de déterminer, à partir d'un plan d'échantillonnage particulier et d'une procédure d'estimation donnée, les chances qu'un certain intervalle construit autour de l'estimé obtenu, contienne la valeur (inconnue) que l'on veut estimer pour la population entière. Le principal objet de l'estimation est de réduire au maximum l'étendue de ces intervalles de façon à ce que les hypothèses sur les valeurs inconnues de la population soient aussi précises que possible. La mesure courante de précision utilisée pour comparer les procédures d'estimation est appelée l'erreur type. En plus de minimiser l'erreur type, le deuxième objectif du choix d'une technique d'estimation de l'échantillon du recensement est d'assurer autant que possible que les estimations des données recensées de l'échantillon soient compatibles avec les valeurs connues de la population correspondante. Par bonheur, ces deux objectifs sont généralement complémentaires, c'est-à-dire que l'erreur d'échantillonnage est habituellement réduite lorsqu'on prend soin que certaines données de base soient compatibles avec les chiffres correspondants de la population. Par contre, le fait de forcer les estimations des données à 100 % de l'échantillon à être compatibles avec les chiffres correspondants de la population de très petites cellules peut avoir de graves conséquences sur l'erreur due à l'échantillonnage des estimations des caractéristiques à 33 % proprement dites.

Lorsqu'on n'a aucune information sur la population échantillonnée autre que celles qui ont été recueillies pour les unités d'échantillonnage, la procédure d'estimation se limite à pondérer les unités d'échantillonnage inversement à leur probabilité de sélection, c'est-à-dire si toutes les unités d'une province avaient une chance sur 10 d'être choisies, toutes les unités de cette province recevraient un coefficient de pondération de 10. En pratique, cependant, on a généralement quelque connaissance de la population observée, par exemple, de sa taille globale et, éventuellement, de sa ventilation en fonction d'une variable donnée, supposons, la province. Ces données peuvent être utilisées pour améliorer la formule d'estimation de façon à produire des estimations qui se rapprochent davantage de la valeur inconnue de la population. Dans le cas de l'échantillon du recensement, on dispose d'un grand nombre de données très détaillées sur la population échantillonnée, soit les données à 100 %, à tous les niveaux géographiques. D'une part, on peut se servir de cette information pour améliorer les estimations faites à partir de l'échantillon de recensement mais d'autre part, cette multitude de données peut également occasionner des ennuis compte tenu du fait qu'il est impossible d'obtenir des estimations de l'échantillon qui soient compatibles avec toutes les données sur la population à tous les niveaux géographiques. Les écarts entre les estimations de l'échantillon et les valeurs de la population apparaissent quand on produit un recoupement d'une variable à 100 % et d'une variable à 33 %. La totalisation doit être fondée sur les données échantillonnées; de ce fait, les totaux marginaux des variables de base sont des

ferent tabulation based on 100% data. They will not necessarily agree exactly. For the sake of brevity, we shall refer to the level of agreement between a sample estimate and the corresponding population figure as "sample population consistency".

3.3 Developing an Estimation Procedure for the Census Sample

Given that a weight has to be assigned to each unit (person, family or household) in the sample, the simplest procedure would be to give each unit a weight of 3 (because a 1 in 3 sample was selected). Such a procedure would be simple and unbiased⁵ and, if nothing but the sample data were known, it might be the optimum procedure. However, although we know that the sample will contain almost exactly one third of all households (excluding collective households and those in canvasser areas), one cannot be certain that it will contain exactly one third of all persons, or one-third of each type of household, or one third of all females aged 25 - 34, and so on. Therefore, this procedure would not ensure consistency even for the most important subgroups of the population. For large subgroups, these fractions should be very close to one third, but for smaller subgroups they could differ markedly from one third. The next most simple procedure would be to define certain important subgroups (e.g., age-sex groups within province) and, for each subgroup, to count the number of units in the population in the subgroup (N) and the number in the sample (n) and to assign to each sample unit in the subgroup a weight equal to N/n .

For example, if there were 5,000 males aged 20 - 24 enumerated in Prince Edward Island, and 1,700 of these fell in the sample households, then a weight of $5,000/1,700 = 2.94$ would be assigned to each male aged 20 - 24 in the sample in Prince Edward Island. This would ensure that whenever sex and age in five-year groups were cross-classified against a sample characteristic for Prince Edward Island, the marginal total for the male 20 - 24 age-sex group would agree with the population total of 5,000. Note that a weight of 3 in this case would result in a sample estimate of 5,100 ($1,700 \times 3$).

⁵ "Unbiased" means that the average, over all possible samples of the estimates obtained by this procedure, would equal the true population value.

estimations échantillonnelles que l'on peut comparer avec les chiffres correspondants de la population d'une autre totalisation basée sur les données à 100 %. Il se peut qu'ils ne concordent pas exactement. Pour simplifier, le degré de concordance entre une estimation d'échantillon et la population correspondante sera désigné sous le nom de "compatibilité entre l'échantillon et la population".

3.3 Élaboration de méthodes d'estimation de l'échantillon du recensement

Étant donné qu'il faut attribuer un coefficient de pondération à chaque unité (personne, famille ou ménage) dans l'échantillon, le plus simple serait de choisir le coefficient 3, puisqu'il s'agit d'un échantillon de 1 sur 3. Une telle méthode serait simple et sans biais⁵, et si l'on n'avait aucune autre donnée que celles de l'échantillon, ce serait la meilleure méthode. Or, bien qu'on sache que l'échantillon contiendra presque exactement le tiers de tous les ménages (à l'exclusion des ménages collectifs et de ceux qui se trouvent dans les secteurs dénombrés par un représentant), on ne peut pas être certain qu'il contiendra exactement un tiers de toutes les personnes, ou un tiers de chaque type de ménage, ou un tiers de toutes les femmes âgées de 25 - 34 ans, etc. Par conséquent, une telle méthode n'assurerait pas la compatibilité, même s'il s'agit des sous-groupes les plus importants de la population. Dans ce dernier cas, les fractions seraient très proches d'un tiers, mais elles pourraient varier de façon sensible dans le cas des petits sous-groupes. Une autre méthode simple serait de prendre certains sous-groupes fondamentaux, par exemple, les groupes selon l'âge et le sexe par province, puis, pour chacun, de compter le nombre d'éléments de la population appartenant au sous-groupe (N) et le nombre d'éléments de l'échantillon appartenant au sous-groupe (n), et d'attribuer à chaque élément d'échantillon un coefficient de pondération égal à N/n .

Par exemple, si l'on recensait 5,000 hommes âgés de 20 - 24 ans dans l'Île-du-Prince-Édouard et si 1,700 de ces hommes faisaient partie des ménages inclus dans l'échantillon, un coefficient de pondération de $5,000/1,700 = 2.94$ serait attribué à chaque homme âgé de 20 - 24 ans dans l'échantillon de l'Île-du-Prince-Édouard. De cette façon, chaque fois que le groupe d'âge quinquennal serait recoupé selon l'âge et le sexe avec une caractéristique de l'échantillon de l'Île-du-Prince-Édouard, le total marginal de la caractéristique de l'âge et du sexe du groupe d'hommes âgés de 20 - 24 ans concorderait avec le total de la population, soit 5,000. À noter que, dans ce cas, un coefficient de pondération de 3 donnerait une estimation de l'échantillon de 5,100 ($1,700 \times 3$).

⁵ "Sans biais" veut dire que la moyenne des estimés de l'ensemble de tous les échantillons possibles serait égale à la valeur de la population réelle.

To understand the type of sample estimate this procedure would produce, consider the tabulation of age by sex by mobility status for Prince Edward Island.

Age-sex	Mobility status		Total
	Migrant	Non-migrant	
Male 20 - 24	X	Y	Z

The estimate, X, is obtained by summing the weights of all males aged 20 - 24 in the sample who are coded as migrant. Thus:

$$X = \text{Weight for males 20 - 24 in P.E.I.} \times \text{Number of male 20-24 migrants in sample in P.E.I.}$$

$$= \frac{\text{Number of male 20 - 24 migrants in P.E.I. in sample}}{\text{Number of males 20 - 24 in P.E.I. in sample}} \times \text{Population of males 20 - 24 in P.E.I.}$$

$$= \text{Proportion of migrants among males 20 - 24 in the sample} \times \text{Population of males 20 - 24 in P.E.I.}$$

Note that the estimate Z will equal the population of males 20 - 24 in Prince Edward Island.

This type of estimation procedure is known as "ratio estimation". It can be shown that this procedure can lead to substantial reductions in standard error in many situations.

This procedure will ensure consistency between sample estimates and population figures for the chosen subgroups and for combinations of these subgroups. However, although the procedure will tend to improve consistency for smaller subgroups (e.g., counties, or single years of age), it will not ensure consistency for these smaller groups, nor for groups defined in terms of other basic characteristics (e.g., marital status, mother tongue). One might consider then extending this procedure to smaller subgroups defined as the cells in a cross-classification of all relevant basic (100%) characteristics. The problem is that, as the subgroup becomes smaller, this procedure becomes unstable (i.e.,

Pour mieux comprendre le type d'estimation échantillonnale produite par cette méthode, prenons comme exemple la totalisation du statut de mobilité dans l'Île-du-Prince-Édouard selon l'âge et le sexe:

Âge-sexe	Statut de mobilité		Total
	Migrant	Non-migrant	
Hommes âgés de 20 - 24 ans	X	Y	Z

La valeur X est obtenue en faisant la somme des coefficients de pondération de tous les hommes âgés de 20 - 24 ans dans l'échantillon qui sont codés comme migrants. Ainsi:

$$X = \text{Le coefficient de pondération des hommes de 20 - 24 ans de l'Île-du-Prince-Édouard} \times \text{Le nombre d'hommes migrants de 20 - 24 ans dans l'échantillon de l'Î. P.-É.}$$

$$= \frac{\text{Le nombre d'hommes migrants de 20 - 24 ans de l'Î. P.-É. dans l'échantillon}}{\text{Le nombre d'hommes de 20 - 24 ans de l'Î. P.-É. dans l'échantillon}} \times \text{La population d'hommes de 20 - 24 ans de l'Î. P.-É.}$$

$$= \text{La proportion de migrants parmi les hommes de 20 - 24 ans de l'Î. P.-É. dans l'échantillon} \times \text{La population d'hommes de 20 - 24 ans de l'Î. P.-É.}$$

À noter que la valeur Z sera égale à la population d'hommes âgés de 20 - 24 ans de l'Île-du-Prince-Édouard.

Cette méthode est connue sous le nom d'"estimation par la méthode du quotient". On peut prouver que, dans de nombreux cas, son utilisation mène à une forte réduction de l'erreur type.

Cette méthode permet d'assurer la compatibilité entre les estimations échantillonnales et les totaux de population des sous-groupes choisis et des combinaisons de ces sous-groupes. Cependant, bien qu'elle permette d'augmenter la compatibilité des plus petits sous-groupes, par exemple, par année d'âge dans les comtés, elle ne permet pas d'assurer la compatibilité de ces groupes ou d'autres groupes définis selon d'autres caractéristiques de base, par exemple, l'état matrimonial ou la langue maternelle. On pourrait envisager d'étendre cette méthode à de plus petits sous-groupes qui seraient définis comme les cellules dans un recoupement de l'ensemble des caractéristiques de base pertinentes. Un problème se pose alors car à

the standard errors of the estimates produced by this procedure increase). In the limit, the procedure becomes impossible when no sample units happen to fall in a particular subgroup. The challenge, therefore, is to obtain the advantages of ratio estimation without suffering the instabilities of using small subgroups. The solution adopted is to carry out ratio estimation iteratively for two distinct and exhaustive sets of subgroups. This procedure, known as the "raking ratio estimation procedure (RREP)", was used in the 1976 Census and is described in the following section.

3.4 The Raking Ratio Estimation Procedure

Instead of just one set of subgroups, two sets of subgroups are defined. One set of subgroups forms the rows of a "weighting matrix" while the other set forms the columns (e.g., for calculating person weights, age-sex-marital status subgroups form the rows of the matrix, while family status-mother tongue subgroups form the columns of the matrix).

Appendix II illustrates two such matrices. Given the appropriate matrix, the RREP proceeds as follows:

- (a) Cross-classify the population records into the matrix to give population totals in each row and column.
- (b) Cross-classify the sample records into the same matrix to give sample totals in each row and column.
- (c) If necessary, collapse the rows and columns of the matrix to meet certain size constraints (see below).
- (d) Calculate the ratio of the number of population records to the number of sample records separately for each collapsed row and column of the matrix.
- (e) Assign an initial weight to each sample record equal to the ratio calculated in (d) for the column in which the record falls. (This will ensure consistency for the columns but not for the rows.)
- (f) Using these initial weights, calculate the estimate for each row.
- (g) For each row, adjust the initial weight of each sample unit in the row by a factor that will ensure that row estimates agree with row population totals. (This will destroy the exact agreement for columns.)
- (h) Repeat the last two steps only this time for columns using the adjusted weight.

mesure que le sous-groupe devient plus petit, cette méthode devient instable, c'est-à-dire que les erreurs types des estimations ainsi produites augmentent. À la limite, cette méthode devient impossible à utiliser lorsqu'un sous-groupe particulier ne comprend plus aucun élément d'échantillon. Le défi consiste donc à profiter des avantages d'une estimation par la méthode du quotient sans subir l'instabilité reliée à l'utilisation de petits sous-groupes. La solution adoptée consiste à appliquer cette méthode de façon itérative à deux ensembles distincts et exhaustifs de sous-groupes. Cette méthode, appelée "méthode itérative d'estimation des quotients" (RREP) et utilisée au recensement de 1976, est décrite ci-après.

3.4 Méthode itérative d'estimation des quotients

Dans cette méthode, on définit deux ensembles de sous-groupes au lieu d'un. Un des ensembles sert à former les lignes d'une "matrice de pondération" tandis que l'autre en forme les colonnes. Ainsi, dans le calcul des coefficients de pondération des personnes, les sous-groupes âge-sexe-état matrimonial forment les lignes de la matrice, tandis que les sous-groupes catégorie de la famille-langue maternelle en forment les colonnes.

L'appendice II donne deux exemples de ce type de matrices. Supposant qu'on ait la matrice appropriée, la RREP se présente comme suit:

- a) On fait le classement recoupé des éléments de la population dans la matrice de façon à obtenir les totaux de la population de chaque ligne et de chaque colonne.
- b) On fait le classement recoupé des éléments de l'échantillon dans la même matrice pour obtenir les totaux d'échantillon de chaque ligne et de chaque colonne.
- c) Si besoin est, on comprime les lignes et les colonnes de la matrice en fonction de certaines limites de taille (voir ci-après).
- d) On calcule séparément le quotient du nombre d'éléments de la population par rapport au nombre d'éléments de l'échantillon de chaque ligne et de chaque colonne de la matrice.
- e) On attribue à chaque élément d'échantillon, dans la colonne où il apparaît, un coefficient de pondération initial égal au quotient calculé en d). Cela permet d'assurer la compatibilité des colonnes mais non des lignes.
- f) On calcule l'estimation de chaque ligne en utilisant ces coefficients initiaux.
- g) À chaque ligne, on ajuste le coefficient initial de chaque élément d'échantillon de la ligne à l'aide d'un facteur permettant d'assurer que les estimations d'une ligne concordent avec les totaux de la population de la même ligne. Cela détruit la concordance exacte des colonnes.
- h) On répète ces deux dernières opérations seulement pour chaque colonne en utilisant le coefficient de pondération ajusté.

- (i) Repeat steps (f) and (g) again for rows using the adjusted weight from step (h).

The procedure stops after four iterations (i.e., column, row, column, row). Row totals will be exactly consistent, and column totals will be approximately consistent. It has been shown that further iterations would not significantly improve the estimates. The important feature of this procedure is that the size limitations (to avoid instability in the estimators) apply only to the row and column totals and not to the individual cells of the matrix (some of which could be empty).

There are two parameters in the RREP which are crucial to the question of consistency between sample estimates for basic characteristics and the corresponding population figures. The first is the choice of the geographic area or weighting area (WA) within which the above procedure is applied. Steps (a) to (i) described above are applied independently within each WA. The second is the choice of the subgroups to define the rows and columns of the weighting matrix.

The WA is the geographic area for which almost exact agreement is ensured for total counts of persons and households and for those subgroups defined by the rows and columns of the weighting matrix. From the point of view of consistency for small areas, the smaller the WA the better. However, the smaller the WA the less detail is possible in the rows and columns of the weighting matrix (because of minimum size limits on these rows and columns). The compromise that was adopted for the 1976 Census was the following:

- (a) a WA should contain between 3,000 and 7,000 persons;
- (b) WA boundaries should respect the boundaries of census divisions (CDs), and as far as possible, of census subdivisions (CSDs), and of census tracts (CTs) where these exist;
- (c) WAs should be made up of whole EAs and should generally be connected.

There are two criteria for choosing subgroups to use in the weighting matrix. First, correlation between the variables defining the subgroups and the sample characteristics is important in minimizing the sampling error of the sample estimates for education, labour force and migration categories. Secondly, the need to ensure consistency for certain important subgroups will influence the choice of rows and columns. These two criteria are often (but not necessarily) complementary. Because of the size limitations on rows and columns (a minimum population of 35 for each row and column was required in 1976), the matrix cannot be too

- i) On répète les opérations f) et g) pour chaque ligne en utilisant le coefficient ajusté à l'étape h).

On s'arrête après quatre itérations, c'est-à-dire colonne, ligne, colonne, ligne. Les totaux des lignes seront alors exactement compatibles, et ceux des colonnes le seront approximativement. Il a été prouvé que les estimations ne seraient pas beaucoup meilleures si l'on effectuait un plus grand nombre de répétitions. Le point important de cette méthode est que les limites de taille visant à éviter l'instabilité des estimateurs s'appliquent seulement aux totaux des lignes et des colonnes et non aux cellules particulières de la matrice, certaines d'entre elles pouvant d'ailleurs être vides.

La RREP comporte deux paramètres qui ont une importance cruciale par rapport à la compatibilité entre les estimations échantillonnelles de caractéristiques à 100 % et les totaux correspondants de la population. Le premier est le choix de la région géographique ou région de pondération (RP) à laquelle on applique cette méthode. Les étapes a) à i) décrites précédemment sont appliquées indépendamment à l'intérieur de chaque RP. Le deuxième est le choix des sous-groupes qui servent à former les lignes et les colonnes de la matrice de pondération.

La RP est cette région géographique pour laquelle une concordance presque exacte est assurée des totaux de personnes et de ménages ainsi que des sous-groupes formant les lignes et les colonnes de la matrice de pondération. Du point de vue compatibilité, plus la RP est petite, plus la compatibilité sera grande. Dans ce cas, cependant, le nombre de détails dans les lignes et les colonnes sera également réduit à cause des limites de taille minimales de ces lignes et de ces colonnes. Lors du recensement de 1976, on a adopté le compromis suivant:

- a) une RP ne devrait contenir que 3,000 à 7,000 personnes;
- b) les limites des RP devraient respecter les limites des divisions de recensement (DR) et, dans la mesure du possible, celles des subdivisions de recensement (SDR) ainsi que celles des secteurs de recensement (SR) lorsqu'il y en avait;
- c) Les RP devraient être composées de SD complets et devraient, de façon générale, être contiguës.

Il y a deux critères pour la sélection des sous-groupes d'une matrice de pondération. En premier lieu, la corrélation entre les variables qui définissent les sous-groupes et les caractéristiques échantillonnelles est importante pour minimiser l'erreur échantillonnelle des estimations échantillonnelles des catégories de la scolarité, de l'activité au travail et de la migration. En second lieu, le besoin de la compatibilité de certains sous-groupes importants influe sur le choix des lignes et des colonnes. Ces deux critères sont souvent, mais pas nécessairement, complémentaires. La matrice ne peut pas être très détaillée à cause des limites de taille des lignes et des

detailed. Rows and columns with a population less than 35 were automatically combined with other rows and columns prior to calculating weights. Two different matrices were used in 1976, one for person and family weights, the other for household weights. The two matrices are shown in Appendix II.

Associated with each matrix was a collapsing strategy that defined how rows and columns were to be combined in any WA in which a row or column had a population less than 35 or a ratio (as calculated in step (d) above) not between 1 and 12. The choice of collapsing strategies was designed to preserve sub-groups wherever possible according to the criteria described in the previous paragraph. In general terms, these collapsing strategies can be summarized as follows:

1. Population and family matrix

- (a) Rows Collapse over age groups within the age groups 0 - 14, 15 - 24 and 25 - 64. For 65 +, collapse over marital status.
- (b) Columns Collapse first over mother tongue. In addition, collapse over "age of children" for spouses in husband-wife families, and collapse the non-family members categories, if necessary.

2. Housing and household matrix

- (a) Rows For family households, collapse first over number of persons and then over sex. For one-person households, collapse first over sex and then over age. For two or more person non-family households, collapse over sex.

- (b) Columns Collapse over type of dwelling.

Collapsing was carried out independently within each WA and ended as soon as all row and column population and ratio constraints were satisfied.

The RREP resulted in a final weight that was the same for all units in the same cell of the collapsed matrix but which differed from cell to cell. These final weights were then added to the record of each sample

colonnes; en 1976, la population minimale requise était de 35 pour chaque ligne et chaque colonne. Les lignes et colonnes dont la population était inférieure à 35 étaient automatiquement agrégées à d'autres lignes et colonnes avant le calcul des coefficients de pondération. En 1976, on a utilisé deux matrices distinctes, une pour les coefficients des personnes et des familles et l'autre, pour les coefficients des ménages. Ces deux matrices sont illustrées à l'appendice II.

Chaque matrice avait sa propre méthode de compression des lignes et des colonnes définissant comment lignes et colonnes devaient être agrégées, et ce, pour chaque RP pour laquelle la population d'une ligne ou d'une colonne était inférieure à 35 ou dont le quotient (voir l'étape d)) n'était pas compris entre 1 et 12. Le choix de la méthode de compression était fait de manière à manipuler les sous-groupes en tenant compte autant que possible des deux critères du paragraphe précédent. En termes généraux, la compression se présentait de la façon suivante:

1. Matrice de la population et des familles

- a) Lignes Compression des groupes d'âge à l'intérieur des groupes: 0 - 14 ans, 15 - 24 ans et 25 - 64 ans. Pour la catégorie d'âge de 65 ans et plus, la compression se fait au niveau de l'état matrimonial.
- b) Colonnes En premier lieu, la compression se fait au niveau de la langue maternelle. Ensuite, elle se fait au niveau de "l'âge des enfants" dans le cas des épouses des familles époux-épouse, puis au niveau des catégories des personnes hors famille, si besoin en est.

2. Matrice du logement et des familles

- a) Lignes Dans le cas des ménages familiaux, compression d'abord selon le nombre de personnes, puis selon le sexe. Dans le cas des ménages d'une personne la compression se fait d'abord au niveau du sexe, puis de l'âge. Dans le cas des ménages non familiaux de deux personnes ou plus, la compression se fait au niveau du sexe.
- b) Colonnes Compression au niveau du type de logement.

La compression a été effectuée séparément pour chaque RP et a été terminée aussitôt que toutes les contraintes s'appliquant à la population des lignes et des colonnes et aux valeurs des quotients ont été satisfaites.

La RREP a produit un coefficient de pondération final qui était le même pour toutes les unités d'une même cellule de la matrice comprimée mais qui variait de cellule à cellule. Ces coefficients finals ont alors été ajoutés aux

unit on the data base. Each person in the sample received the weight calculated for the cell of the population and family matrix in which he/she fell; each household in the sample received the weight from the appropriate cell of the housing and household matrix; each family in the sample received the personal weight of a designated adult family member. Persons, households and families in those sectors of the population enumerated on a 100% basis automatically received a weight equal to one.

Operationally, the RREP was almost fully automated. Weighting areas were formed using a program that takes into account EA population, geographic co-ordinates of EA centroids, and the geostatistical areas (CDs, CSDs, etc.) in which the EA is located. This program provided a listing of the WAs thus formed and allowed changes to be made manually if appropriate. This facility for manual adjustment was used in a small number of cases. Once the WAs had been fixed the cross-classification of data, the collapsing of rows and columns, the calculation of weights, and the assignment of these weights to records on the data base took place entirely within the computer.

enregistrements de chaque unité de l'échantillon dans la base des données. À chaque personne dans l'échantillon, on a attribué le coefficient de pondération calculé pour la cellule de la matrice de la population et des familles à laquelle elle appartenait; à chaque ménage dans l'échantillon, on a attribué le coefficient de pondération de la cellule appropriée de la matrice du logement et des ménages, tandis qu'on a attribué à chaque famille dans l'échantillon le coefficient de pondération de l'un de ses membres adultes désigné à l'avance. Un coefficient de pondération égal à un a été automatiquement attribué à toutes les personnes, toutes les familles et tous les ménages appartenant à des secteurs recensés.

Physiquement, la RREP était presque entièrement automatisée. Les RP furent formées en utilisant un programme qui tenait compte de la population du SD, des coordonnées géographiques des centroïdes des SD et des secteurs géostatistiques (DR, SDR, etc.) auxquels les SD appartiennent. Ce programme a fourni un output énumérant les RP créées et permettait des modifications manuelles si nécessaire. On ne fit de modifications manuelles que dans un nombre restreint de cas. Une fois la détermination des RP terminée, le recoupement des données, la compression des lignes et colonnes, le calcul des poids et leur insertion dans les enregistrements de la base des données fut entièrement réalisé par l'ordinateur.

IV. THE SAMPLING AND WEIGHTING EVALUATION PROGRAM

The sampling and weighting evaluation program was designed to answer a number of questions related to the effect of sampling and weighting on the quality of census sample data. Sampling and weighting of course have no effect on 100% census data. The principal questions were:

1. Were the sample selection procedures correctly implemented, and was the resulting sample equivalent to a simple random sample of private households within each EA?
2. Were the various stages of the weighting procedures correctly implemented and was the resulting degree of consistency between sample estimates and 100% counts satisfactory?
3. What was the contribution of sampling error to the overall quality of census sample data?

To answer the first two questions, and to provide other information relevant to the planning of future censuses, four studies were carried out within the sampling and weighting evaluation program. These studies were:

- (a) check on drop-off bias;
- (b) examination of sample population consistency;
- (c) evaluation of procedures for forming WAs;
- (d) evaluation of cross-classification and collapsing procedures.

To answer the third question, information on the contribution of sampling variance to total variance for census sample data was obtained from the *Total Variance Study*.⁶

In the remainder of this chapter, the four studies listed above are briefly described. Chapters V, VI and VII are devoted to the presentation of information in response to Questions 1, 2 and 3, respectively.

(a) Drop-off bias check

The objective of this study was to determine whether the sample selection procedures had been correctly implemented. This question was approached from two angles. First, an examination of the routes taken by a sample of CRs was carried out to spot

⁶ Total variance is defined in Chapter VII. For more information on the *Total Variance Study*, see Brackstone and Hill, 1976.

IV. PROGRAMME D'ÉVALUATION DE L'ÉCHANTILLONNAGE ET DE LA PONDÉRATION

Le programme d'évaluation de l'échantillonnage et de la pondération a été conçu en vue de répondre à un certain nombre de questions en rapport avec l'incidence des méthodes d'échantillonnage et de pondération sur la qualité des données du recensement. Évidemment, l'échantillonnage et la pondération n'ont pas d'effet sur les données recensées. Les principales questions que l'on s'est posé étaient les suivantes:

1. Les méthodes de sélection de l'échantillon ont-elles été correctement appliquées et l'échantillon obtenu est-il équivalent à un échantillon aléatoire simple de ménages privés pris à l'intérieur de chaque SD?
2. Les diverses étapes de la procédure de pondération ont-elles été correctement appliquées et les estimations faites à partir de l'échantillon et des chiffres se rapportant à l'ensemble de la population sont-elles suffisamment compatibles?
3. Quelle est la contribution de l'erreur échantillonnale à la qualité globale des données de l'échantillon du recensement?

Afin de répondre aux deux premières questions et d'obtenir d'autres informations reliées à la planification de recensements futurs, on a procédé, dans le cadre du programme d'évaluation de l'échantillonnage et de la pondération, à quatre études qui avaient pour objet:

- a) le contrôle des erreurs à la livraison;
- b) l'étude de la compatibilité des données de l'échantillon et de celles de l'ensemble de la population;
- c) l'évaluation de la méthode de formation des régions de pondération (RP);
- d) l'évaluation des procédures de recoupement et d'agrégation.

Pour répondre à la troisième question, on a eu recours à l'*Étude de la variance totale*⁶ pour établir l'incidence de la variance échantillonnale sur la variance totale des données de l'échantillon du recensement.

La fin du présent chapitre donne un exposé sommaire des quatre études citées ci-dessus. Les chapitres V, VI et VII présentent les données qui répondent aux questions 1, 2 et 3, respectivement.

a) Contrôle des erreurs à la livraison

Cette étude visait à déterminer si les méthodes de sélection de l'échantillon avaient été correctement appliquées. La question a été abordée sous deux angles. Tout d'abord, on a procédé à une étude des itinéraires suivis par un échantillon de RR, afin de déterminer si les procédures

⁶ On trouvera la définition de la variance totale au chapitre VII. Pour plus de détails concernant l'*Étude de la variance totale*, voir Brackstone et Hill, 1976.

deviations from the prescribed procedures. Secondly, for the same sample of EAs, the characteristics of the sample households were compared with those of all households to see whether there were any statistically significant differences. Finally, the relationship between apparent deviations from the correct sampling procedures and the existence of statistically significant differences at the EA level was examined.

(b) Sample population consistency

The objective of this study was to measure and summarize the extent to which agreement between sample estimates and population values had been achieved. The level of agreement was examined for a variety of geographic areas and for a variety of 100% categories. For that part of the population that was sampled, a comparison was also made with the crude estimate obtained by multiplying the sample count by three.

(c) Evaluation of procedures for forming WAs

The objective of this study was to measure the degree to which WAs met the criteria laid down for their formation (see Section 3.4). For this purpose, a sample of census divisions⁷ was selected and the respect for boundaries, the sizes and the contiguity of WAs formed within those CDs were examined and analysed.

(d) Cross-classification and collapsing

This study measured to what extent important categories of persons and households were preserved after collapsing. It produced frequency counts of the number of WAs in which various levels of collapsing were necessary. It also attempted to identify changes to the cross-classification and collapsing strategies that would make the weighting system more effective in future censuses.

⁷ Census division (CD) is a general term applying to counties, regional districts, regional municipalities, etc. In Newfoundland, Manitoba, Saskatchewan and Alberta, the term describes geostatistical areas that have been created by Statistics Canada in co-operation with the provinces as an equivalent for counties.

établies avaient été respectées. Dans un deuxième temps, on a comparé les caractéristiques des ménages échantillonnés dans un même SD, avec celles de l'ensemble des ménages pour voir s'il y avait des différences significatives du point de vue statistique. On a ensuite vérifié s'il y avait un rapport entre les écarts attribuables au fait que les méthodes d'échantillonnage n'avaient pas été suivies et la présence de différences significatives du point de vue statistique au niveau des SD.

b) Compatibilité des données de l'échantillon et de l'ensemble de la population

Cette étude avait pour objet de mesurer le degré de concordance des estimations faites à partir de l'échantillon et des totaux de l'ensemble de la population pour un certain nombre de régions géographiques et de catégories à 100 %. On a également établi une comparaison entre la population échantillonnée et les estimations brutes obtenues en multipliant par trois les chiffres de l'échantillon.

c) Évaluation de la méthode de formation des régions de pondération

Cette étude visait à déterminer dans quelle mesure les régions de pondération répondaient aux critères prévus pour leur établissement (voir le paragraphe 3.4). À cette fin, on a prélevé un échantillon de divisions de recensement⁷, puis on a analysé les régions de pondération créées à l'intérieur de ces DR, du point de vue des limites, de la taille et de la contiguïté.

d) Recoupement et agrégation

Cette étude avait pour objet de mesurer jusqu'à quel point les catégories importantes de personnes et de ménages étaient touchées par l'agrégation. Elle a permis d'établir les fréquences de régions de pondération ayant nécessité divers niveaux d'agrégation. Elle a également visé à déterminer quelles modifications des procédures de recoupement et d'agrégation seraient souhaitables afin d'améliorer la pondération pour les recensements futurs.

⁷ L'expression "division de recensement" (DR) est un terme général qui désigne les comtés, les districts régionaux, les municipalités régionales, etc. À Terre-Neuve, au Manitoba, en Saskatchewan et en Alberta, l'expression désigne les secteurs géostatistiques créés comme équivalents des comtés par Statistique Canada en consultation avec les provinces.

V. EVALUATION OF SAMPLE SELECTION

The procedures for selecting the sample within each EA were described in Section 2.3. If correctly implemented, these procedures result in a systematic sample of households stratified by EA. The principal reason for utilizing systematic sampling is the operational simplicity of the procedure. However, in addition, since the systematic procedure ensures that households from all parts of the EA are represented in the sample, it will often be superior to simple random sampling (i.e., sampling in which every possible sample of a given size has the same chance of being selected) in terms of its average ability to reflect accurately the characteristics of all households in an EA. Simple random sampling of households will be used as a benchmark for comparison in the sense that we wish to check that the systematic sampling procedure as implemented leads to samples of households that, on average, represent all households at least as well as would have been expected under a simple random sampling procedure.

Since the sampling procedures were implemented separately by more than 30,000 CRs across Canada, the possibility of deviation from the correct sampling procedures in some cases is very real. The main concern is not with the occasional random error that CRs might make but with any systematic deviation from procedures that might be perpetrated by many CRs. If, for example, the work involved in obtaining a completed long form is seen by CRs to be significantly more than for the short form, there may be a common temptation to avoid leaving a long form with a household that appears more difficult to enumerate (e.g., a large household, or a household unable to speak English or French). Any such systematic tampering with the sampling procedures has the potential for introducing bias into the sample.

The purpose of this evaluation was, first, to see whether the CRs had correctly implemented the sampling procedures, secondly, to see whether the correspondence between the characteristics of sample households and those of all households was at least as close as would be expected under simple random sampling, and thirdly to see whether there was any association between a CR's failure to follow the correct sampling procedures and unrepresentativeness in the resulting sample.

5.1 Evaluation of the CR's Route

A CR's adherence to sampling procedures was judged by examining the route taken around his/her EA and by the order of listing of dwellings in the VR. To conduct this examination (and to conduct the tests described in Section 5.2), a sample of 1,000 EAs was selected from all EAs in which sampling was used in the 1976 Census. The sample of

V. ÉVALUATION DE LA MÉTHODE DE SÉLECTION DE L'ÉCHANTILLON

Les techniques de sélection de l'échantillon dans chaque SD ont été décrites au paragraphe 2.3. Une application fidèle de ces techniques donne un échantillon systématique de ménages stratifiés par SD. La simplicité d'application de la méthode de l'échantillon systématique est la principale raison de son utilisation. Toutefois, étant donné qu'en plus, l'échantillon systématique contient inévitablement des ménages de toutes les parties du SD, cette méthode est souvent supérieure à celle de l'échantillonnage aléatoire simple (méthode selon laquelle tous les échantillons possibles d'une taille donnée ont les mêmes chances d'être sélectionnés) parce qu'elle permet en moyenne de tenir compte de façon précise des caractéristiques de tous les ménages d'un SD. L'échantillonnage aléatoire simple des ménages sera utilisé comme étalon pour vérifier que la procédure d'échantillonnage systématique a donné lieu à des échantillons de ménages qui, en moyenne, représentent l'ensemble des ménages au moins aussi bien que les échantillons de ménages résultant d'un échantillonnage aléatoire simple.

Étant donné que plus de 30,000 RR ont eu à appliquer individuellement les procédures d'échantillonnage dans tout le Canada, il est inévitable que des écarts se soient produits. L'erreur la plus redoutée n'est pas l'erreur aléatoire accidentelle que tout RR peut commettre, mais plutôt une mauvaise application systématique des procédures de la part de quelques-uns d'entre eux. Ainsi, un RR qui considère que les questionnaires complets représentent un surcroît de travail par rapport aux questionnaires abrégés pourrait être tenté de ne pas remettre un questionnaire complet aux ménages difficiles à dénombrer, par exemple, ceux qui comportent un grand nombre de membres ou ceux dont les membres ne parlent ni anglais ni français. De telles déviations des méthodes d'échantillonnage risquent de biaiser l'échantillon si le comportement en question se répète trop souvent.

L'évaluation consistait, premièrement, à s'assurer que les RR ont correctement appliqué les procédures d'échantillonnage, deuxièmement, à vérifier que la correspondance entre les caractéristiques des ménages de l'échantillon et celles de l'ensemble des ménages est au moins aussi précise que celle obtenue dans le cas d'un échantillonnage aléatoire simple, et, troisièmement, à voir s'il y a un rapport entre une mauvaise application des méthodes d'échantillonnage par un RR et le manque de représentativité de l'échantillon résultant.

5.1 Évaluation de l'itinéraire du RR

Pour évaluer la mesure dans laquelle un RR a respecté les procédures d'échantillonnage, on a examiné l'itinéraire qu'il a suivi dans son SD ainsi que l'ordre dans lequel il a dressé la liste des logements sur le RV. À cette fin, et également en vue des tests décrits au paragraphe 5.2, on a pris un échantillon d'environ 1,000 SD dans l'ensemble des SD ayant fait l'objet d'échantillonnage au

EAs was selected with equal probability within each of 20 strata formed by splitting the EAs in each province into those enumerated using mail-back procedures and those enumerated using pick-up procedures. The number of EAs selected in each stratum is given in Table 1.

recensement de 1976. L'échantillon a été choisi avec la même probabilité dans chacune des 20 strates obtenues en classant les SD de chaque province selon que les questionnaires étaient retournés par la poste ou repris par le RR. Le tableau 1 donne le nombre de SD sélectionnés dans chaque strate.

TABLE 1. Number of EAs Selected in Each Stratum for the Evaluation of Sample Selection Procedures
TABLÉAU 1. Nombre de SD choisis dans chaque strate pour l'évaluation des méthodes de sélection des échantillons

Province	Mail-back — Retour par la poste	Pick-up — Reprise
Newfoundland — Terre-Neuve	50	13
Prince Edward Island — Île-du-Prince-Edouard	15	5
Nova Scotia — Nouvelle-Écosse	50	16
New Brunswick — Nouveau-Brunswick	50	15
Québec	150	60
Ontario	160	78
Manitoba	50	15
Saskatchewan	50	33
Alberta	50	32
British Columbia — Colombie-Britannique	75	33
Canada	700	300

Each EA examined was classified into one of three categories:

- (a) no deviation whatsoever from the prescribed route;
- (b) some deviation from the prescribed route but not of a nature expected to affect the sample selection (e.g., some blocks enumerated anti-clockwise instead of clockwise; some blocks not begun at a corner; some households missed at drop-off);
- (c) more drastic deviations from the prescribed route that might affect the sample selection.

It should be pointed out that the classification of EAs into these categories on the basis of maps and VRs after the census presents many difficulties, particularly in rural areas. There may have been very good explanations for some of the apparent deviations from prescribed routes. In examining maps and VRs in an office after the event, one can only form an incomplete picture of the real situation in the field. Thus, the results of this classification are not a reflection on the performance of individual CRs but rather an indication of whether, in general, there was any association between deviation from the prescribed route and unrepresentativeness of sample households. It should also be stressed that deviation from the prescribed route does not necessarily result in sample selection bias. There may be many reasons for deviations from the route quite unrelated to the characteristics of

Tous les SD ainsi examinés ont été classés en trois catégories:

- a) les méthodes établies ont été correctement suivies;
- b) il y a eu erreur dans l'application des procédures mais il est peu probable que cela influe sur la sélection de l'échantillon; par exemple, le dénombrement des îlots urbains a été fait dans le sens contraire des aiguilles d'une montre, ou n'a pas été fait à partir d'un coin de rue, ou encore certains ménages ont été oubliés au moment de la livraison des questionnaires;
- c) les erreurs commises au niveau de l'application des procédures sont de nature à nuire sérieusement à la sélection de l'échantillon.

Il convient de souligner qu'une fois le recensement terminé, la classification des SD dans l'une de ces catégories à partir des cartes et des RV présente beaucoup de difficultés, particulièrement dans les régions rurales. Il peut y avoir certains cas où l'itinéraire établi n'a pas été suivi pour des raisons très valables. L'examen après coup des cartes et des RV dans un bureau ne donne pas une image complète de la situation réelle sur le terrain. Une telle classification ne peut être le reflet du rendement de chaque RR, mais donne plutôt une indication générale sur un rapport éventuel entre le fait que l'itinéraire prévu n'a pas été suivi et le manque de représentativité des ménages de l'échantillon. Il faut également spécifier qu'une déviation de l'itinéraire prévu n'a pas nécessairement pour effet de biaiser la sélection de l'échantillon. Un écart à l'itinéraire prévu peut se produire pour des raisons sans lien avec les caractéristiques des ménages individuels. Par

individual households. On the other hand, adherence to a predefined route, together with the random start, guarantees no bias in sample selection.

Because of the difficulties with rural EAs, this analysis was carried out only for 696 urban EAs in the sample.⁸ Of these, 92 (13%) fell in category (a), 409 (59%) fell in category (b), and 195 (28%) fell in category (c).

5.2 Comparison of Sample Households With All Households

In comparing the characteristics of sample households with those of all households, one has to recognize that the differences can be expected purely by chance. To examine whether the sample adequately represented all households, a series of statistical tests were performed. These tests were designed to see whether the differences between the characteristics of sample households and the characteristics of all households were any larger than could be expected by chance under simple random sampling. Test 1 compared a mean or proportion among sample households with the corresponding quantity for all households. It was applied for a variety of variables (e.g., mean household size, proportion of owner-occupied households). Test 2 compared the distribution of a variable, category by category, for sample households with the corresponding distribution for all households and was also applied for a variety of variables (e.g., type of dwelling, mother tongue of head of household). The sample of 1,000 EAs described in the previous section was used. Both tests were applied at the EA level and at the stratum level. The list of variables and categories considered for each test is given in Appendix III.

The results of testing sample means and proportions at the EA level produced no evidence of any serious departure from simple random sampling. The proportion of EAs producing significant results at the 5% level⁹ was in fact 5% or less in all but one case. For this case, namely the proportion of households with a head whose mother tongue was English, the probability of obtaining at least the number of significant results actually observed under the hypothesis of simple random sampling was about 45%. Thus, there was no evidence of bias in the sample mean or proportion for any of the variables examined. For the important

contre, le respect de l'itinéraire établi et de l'origine au hasard est une garantie que la sélection de l'échantillon n'est pas biaisée.

En raison des difficultés rencontrées au niveau des SD ruraux, cette analyse n'a porté que sur 696 SD urbains de l'échantillon⁸. Quarante-vingt douze SD (13 %) ont été classés dans la catégorie a), 409 (59 %) dans la catégorie b), et 195 (28 %) dans la catégorie c).

5.2 Comparaison des ménages de l'échantillon avec l'ensemble des ménages

En ce qui a trait à la comparaison des caractéristiques des ménages de l'échantillon avec celles de l'ensemble des ménages, on doit reconnaître que les écarts ne peuvent se produire que de façon aléatoire. On a procédé à une série de tests statistiques pour voir si l'échantillon représentait adéquatement tous les ménages. Ces tests ont été conçus en vue de déterminer si les écarts entre les caractéristiques des ménages de l'échantillon et celles de l'ensemble des ménages sont supérieurs à ceux qui pourraient se produire de façon fortuite dans un échantillon aléatoire simple. Dans le premier test, on a comparé une moyenne ou une proportion des ménages de l'échantillon avec la quantité correspondante de l'ensemble des ménages. On a appliqué ce test à un groupe de variables, par exemple, la taille moyenne des ménages et la proportion de logements occupés par leur propriétaire. Dans le deuxième test, on a comparé la répartition d'une variable donnée, catégorie par catégorie, dans les ménages de l'échantillon avec sa répartition dans l'ensemble des ménages. Ce test a aussi été appliqué à un groupe de variables, comme le type de logement et la langue maternelle du chef de ménage. On a utilisé l'échantillon de SD décrit au paragraphe précédent. Ces deux tests ont été appliqués au niveau des SD et au niveau des strates. L'appendice III donne la liste des variables et des catégories considérées pour chacun des tests.

Les résultats des tests sur l'échantillon ont révélé que les moyennes et les proportions observées ne s'écartaient pas de façon significative de celles d'un échantillon aléatoire simple. La proportion de SD pour lesquels les tests à 5 %⁹ ont donné des résultats significatifs était, de fait, de 5 % ou moins dans tous les cas, sauf un. Dans ce dernier cas, qui portait sur la proportion de ménages ayant un chef de langue maternelle anglaise, la probabilité d'avoir le même nombre de résultats significatifs que dans un échantillon aléatoire simple était de 45 % environ. Par conséquent, on n'a relevé aucun biais dans les moyennes ni dans les proportions de l'échantillon selon les variables

⁸ Information was unavailable for the other four urban EAs in the sample.

⁹ Testing at the 5% level means that when the hypothesis being tested is true, there is a 5% chance of obtaining a significant result and thus wrongly rejecting the hypothesis. Therefore, in repeated independent tests at the 5% level, one expects about 5% of the tests to give significant results even when the hypothesis of simple random sampling is true.

⁸ On ne disposait pas des données se rapportant aux quatre autres SD urbains de l'échantillon.

⁹ Par test à 5 %, on veut dire que lorsque l'hypothèse de départ est vraie, il y a une probabilité de 5 % d'avoir un résultat significatif et, de là, un rejet erroné de l'hypothèse. Par conséquent, si l'on effectue plusieurs fois des test indépendants à 5 %, on peut s'attendre à ce que 5 % des tests donnent des résultats significatifs, même lorsque l'hypothèse de l'échantillon aléatoire simple est vraie.

variable "household size", almost exactly 5% of EAs produced significant results at the 5% level and almost exactly 1% at the 1% level. The fact that many variables produced less significant results at the 5% level than would be expected under simple random sampling suggests that the sampling procedure was producing samples that were on average more representative of all households than would be expected with simple random sampling. This was particularly noticeable for the tenure and type of dwelling variables. For these variables, it appears that the systematic nature of the sample has indeed led to more representative samples than would have occurred with simple random sampling.

When sample means and proportion were tested for each of the 20 strata, two cases showed more than the expected one significant result at the 5% level. These were the mother tongue categories "French" and "Other". The four strata which showed significant results for the "French" category were in Newfoundland, Saskatchewan, Alberta and British Columbia and all were only marginally significant (none was significant at the 1% level). In two of these strata, the "French" mother tongue group was slightly over-represented while in the other two, the reverse was true. In each of the five strata in which the "Other" mother tongue category produced significant results, persons in this category were over-represented in the sample. This phenomenon stems most probably from certain problems experienced in the processing of mother tongue data rather than from the sampling procedures. The "Other" mother tongue category corresponds to the office-coded entries on the questionnaire, i.e., those mother tongues for which no self-coding circle was provided for respondents. In reading the coded areas for mother tongue on short form questionnaires, a problem of picking up codes that were not really there occurred.¹⁰ This problem did not occur on the sample questionnaires. A special computer program was introduced to recognize false pick-ups as far as possible and to delete them. However, since it was not possible to determine precisely in every case whether a particular code was a false pick-up or a valid entry, this program could not eliminate the effect of the problem completely. An evaluation of this problem appears to suggest that the program may have slightly over-compensated with the result that some valid entries were lost. Since the problem was confined to the short questionnaire, the net effect of the problem and its partial solution would be to lead to an understatement of the "Other"

étudiées. En ce qui concerne la taille des ménages, qui est une variable particulièrement importante, presque exactement 5 % des SD ont produit des résultats significatifs dans les tests à 5 % et presque exactement 1 % dans les tests à 1 %. Le fait que, pour un bon nombre des variables, les résultats étaient moins significatifs dans les tests à 5 % qu'ils ne l'auraient été dans un échantillon aléatoire simple indique que la méthode d'échantillonnage utilisée donne en moyenne des échantillons plus représentatifs de l'ensemble des ménages que ne peut faire l'échantillonnage aléatoire simple. Ce fait était particulièrement apparent pour des variables telles que le mode d'occupation et le type de logement puisque c'est surtout dans le cas de ces dernières que le caractère systématique de l'échantillon a eu pour résultat de donner des échantillons plus représentatifs que ceux qu'on aurait obtenus par un échantillonnage aléatoire simple.

Les tests appliqués, dans chacune des 20 strates, aux moyennes et aux proportions de l'échantillon ont été, dans deux cas, plus élevés que le résultat significatif attendu: les catégories langue maternelle "Français" et "Autre". Les quatre strates qui ont présenté des résultats significatifs pour la catégorie "Français" appartenaient à Terre-Neuve, à la Saskatchewan, à l'Alberta et à la Colombie-Britannique et toutes n'avaient qu'une signification marginale (aucune n'était significative à 1 %). Dans deux de ces strates, le groupe de langue maternelle française était représenté de façon légèrement excessive, tandis que l'inverse s'est produit dans deux autres strates. Dans chacune des cinq strates qui avaient produit des résultats significatifs dans la catégorie de langue maternelle "Autre", la représentation dans l'échantillon des personnes appartenant à cette catégorie était excessive. Cela est fort probablement attribuable à certains problèmes rencontrés au niveau du dépouillement des données sur la langue maternelle plutôt qu'aux procédures d'échantillonnage. La catégorie de langue maternelle "Autre" correspond aux inscriptions des questionnaires codés au bureau, c'est-à-dire aux langues maternelles pour lesquelles il n'y avait pas de cercle à noircir par le répondant. Dans le cas du questionnaire abrégé, en parcourant la zone des codes de la catégorie langue maternelle, on se heurtait au problème de choisir des codes qui, en réalité, ne s'y trouvaient pas¹⁰. Ce problème ne s'est pas posé pour le questionnaire-échantillon. On a élaboré un programme informatique visant à dépister les réponses mal codées dans la mesure du possible, et à les annuler. Toutefois, comme il n'était pas toujours possible de déterminer si le code choisi était mauvais ou non, ce programme n'a pas permis d'éliminer totalement les effets du problème. D'après une évaluation du problème, il semble que le nombre des erreurs ainsi corrigées par le programme était quelque peu excessif, ce qui a entraîné la perte de certaines bonnes réponses. Comme ce problème ne touche que les

¹⁰ A similar problem occurred for the relationship to head of household question.

¹⁰ Un problème semblable s'est posé pour la question portant sur le lien avec le chef de ménage.

category on short forms. In a check on sampling bias, this would appear as an apparent over-representation of persons in this category in the sample. Table 2 compares a selection of means and proportions for sample households with the corresponding quantities for all households. These figures are derived from the 1,000 EAs in the sample and provide an indication of the level of consistency between sample figures and 100% figures before the weighting procedures are applied.

questionnaires abrégés, l'effet net de ce problème et de solution partielle serait de sous-estimer la catégorie "Autre" de la langue maternelle des questionnaires abrégés. Dans le cadre d'une évaluation du biais dû à l'échantillonnage, ceci apparaîtrait comme une sur-représentation des personnes de cette catégorie dans l'échantillon. Le tableau 2 compare une sélection de moyennes et de proportions des ménages de l'échantillon avec les quantités correspondantes de l'ensemble des ménages. Ces chiffres ont été calculés à partir des 1,000 SD qui composent l'échantillon et fournissent une indication du niveau de compatibilité des données à 33 % avec les données à 100 % avant l'application des procédures de pondération.

TABLE 2. Means and Proportions Among (a) Private Households in the Sample and (b) All Private Households, for a Random Sample of 1,000 EAs

TABLEAU 2. Moyennes et proportions a) des ménages et des logements privés dans l'échantillon par rapport à b) l'ensemble des ménages et des logements privés, dans un échantillon aléatoire de 1,000 SD

Mean or proportion — Moyenne ou proportion	Value for sample households — Valeur des ménages dans l'échantillon	Value for all households — Valeur de l'ensemble des ménages
<i>δ</i>		
Average household size — Taille moyenne des ménages	3.1057	3.1044
Average age of head of household — Âge moyen du chef de ménage	44.5800	45.3600
Proportion of households that are owner-occupied — Proportion de logements occupés par le propriétaire	0.5788	0.5793
Proportion of households occupying single dwellings — Proportion des ménages occupant des logements individuels	0.5140	0.5151
Proportion of households occupying apartments — Proportion des ménages occupant des appartements	0.3219	0.3216
Proportion of heads of household whose mother tongue is English — Proportion des chefs de ménage dont la langue maternelle est l'anglais.	0.6353	0.6380
Proportion of heads of household whose mother tongue is French — Proportion des chefs de ménage dont la langue maternelle est le français	0.2119	0.2126

The tests of the distributions of sample households against the corresponding distributions among all households at the EA level produced no evidence of any differences larger than would be expected under simple random sampling. Again the distributions of the samples by tenure and by type of dwelling tended to be closer than expected to the corresponding distributions among all households. At the stratum level, this test produced two (out of 20) significant results for average household size, for mother tongue of head, and for average age of head. The chances of obtaining two or more significant results assuming simple random sampling are about 26% and therefore these significant results are not inconsistent with the hypothesis of simple random sampling.

La comparaison des résultats des tests sur la répartition des ménages de l'échantillon et de l'ensemble des ménages au niveau du SD n'a révélé aucun écart supérieur à ceux prévus pour un échantillon aléatoire simple. De même, la répartition des ménages de l'échantillon selon le mode d'occupation et le type de logement semblait concorder davantage que prévu avec les répartitions correspondantes de l'ensemble des ménages. Au niveau des strates, ce test a donné deux résultats significatifs (sur 20) pour les catégories "taille moyenne des ménages", "langue maternelle du chef" et "âge moyen du chef". La probabilité d'avoir deux résultats significatifs ou plus dans un échantillon aléatoire simple est d'environ 26 % et, par conséquent, les résultats significatifs obtenus ne sont pas incompatibles avec l'hypothèse de l'échantillon aléatoire simple.

5.3 Analysis and Summary

To investigate whether there was any association between a CR's apparent failure to adhere to the sampling procedures and unrepresentativeness in the resulting sample, a cross-tabulation of EA by route classification and by test results (significant at the 5% level in at least one test, or not significant at the 5% level in any test) was created for the variables: household size; age of head of household; mother tongue of head of household; type of dwelling. These cross-classifications are summarized in Table 3.

5.3 Analyse et résumé

En vue de déterminer s'il y a un rapport entre les cas où le RR n'a apparemment pas respecté les procédures d'échantillonnage et le manque de représentativité de l'échantillon résultant, on a procédé à un classement recoupé des SD selon les itinéraires et selon les résultats des tests (résultats significatifs à 5 % pour au moins un test et non significatifs à 5 % pour tous les tests), selon les variables suivantes: taille des ménages, âge du chef de ménage, langue maternelle du chef de ménage, type de logement. Les classements obtenus sont présentés au tableau 3.

TABLE 3. Number (and Percentage) of EAs Having at Least One Significant Test Result at the 5% Level by Variable and by Route Classification (Based on 696 Urban EAs)

TABLEAU 3. Nombre (et pourcentage) de SD pour lesquels on a obtenu au moins un résultat significatif du test à 5 % selon certaines variables et selon la catégorie d'itinéraire (Basé sur 696 SD urbains)

Variable	Route classification Catégorie d'itinéraire			All EAs Ensemble des SD
	(a)	(b)	(c)	
Household size – Taille du ménage	7 (8%)	30 (7%)	13 (7%)	50 (7%)
Age of head – Âge du chef	3 (3%)	22 (5%)	11 (6%)	36 (5%)
Mother tongue of head – Langue maternelle du chef. .	12 (13%)	76 (19%)	25 (13%)	113 (16%)
Type of dwelling – Type de logement	2 (2%)	13 (3%)	12 (6%)	27 (4%)
Total number of EAs – Nombre total de SD	92	409	195	696

Note: It is the differences between percentages for the three route classifications that are of relevance in this table and not the differences between variables. The differences in percentages between variables result in a large part from the different number of tests applied for each variable. – Nota: Dans ce tableau ce qu'il est pertinent de regarder est l'écart de pourcentages entre les trois catégories d'itinéraire et non l'écart entre les variables. Les écarts de pourcentage entre les variables sont en grande partie attribuables aux nombres différents de tests appliqués à chaque variable.

A statistical test of the differences between these proportions for the three route classifications showed that none of these differences are significant even at the 10% level. Thus despite the fairly widespread incidence of "incorrect" route classifications, there appears to be no association between this and unrepresentativeness of the sample households.

In summary, these studies have not revealed any evidence of serious lack of equivalence to simple random sampling and households in the implementation of the 1976 Census sampling procedures in terms of the variables examined here. In fact, it appears, particularly for housing variables, that the systematic nature of the sampling procedure leads to a more representative sample than would be expected under

On a soumis les différences relevées entre ces proportions pour chacun des trois types d'itinéraires à un test statistique et aucun n'est apparu significatif, même au niveau de 10 %. Il semble donc que malgré le nombre important d'itinéraires "incorrects", il n'y a aucun rapport entre eux et le manque de représentativité des ménages de l'échantillon.

En résumé, les études faites n'ont révélé aucune incompatibilité sérieuse entre les résultats des procédures d'échantillonnage appliquées au recensement de 1976 et ceux d'un échantillon aléatoire simple par rapport aux variables étudiées. En fait, à cause de l'aspect systématique du type d'échantillonnage appliqué, l'échantillon obtenu est plus représentatif qu'un échantillon aléatoire simple, particulièrement en ce qui concerne les variables

simple random sampling. Although this study has not revealed any pervasive sampling bias, there will undoubtedly be individual EAs in which by chance the sample households do not well represent all households. Such deviations will tend to cancel out over many EAs and will in any case be dampened to some extent by the weighting procedures which attempt to pull the weighted sample counts into line with the counts from all households.

des ménages. Bien que l'étude n'ait révélé aucun biais d'échantillonnage systématisé, il peut se trouver des cas où les ménages de l'échantillon ne sont pas vraiment représentatifs de tous les ménages du SD. Ces cas se trouveraient noyés dans l'ensemble des SD et, de toute façon, les données correspondantes seraient corrigées dans une certaine mesure au moment de la pondération qui vise à aligner les chiffres de l'échantillon sur ceux de l'ensemble des ménages.

VI. EVALUATION OF WEIGHTING PROCEDURES AND SAMPLE POPULATION CONSISTENCY

6.1 Weighting Area Formation

The first stage of the weighting procedures was the formation of WAs (see Section 3.4). The sampled EAs were formed into 4,798 WAs with an average population (excluding persons in collective dwellings) of 4,516. Of the 4,798 WAs, 4,672 (97.5%) initially fell within the population limits of 3,000 - 7,000. Of the remaining 126 WAs, 98 were manually adjusted to bring them within these limits. The remaining 28 WAs all had a population between 2,500 and 3,000 but were left as they were in order not to further violate CT or CSD boundaries.

The connectedness of WAs was examined only for rural areas since in urban areas, where WAs are based on compact CTs, no serious problems of connectedness arise. In rural areas where CSDs may be large and irregularly shaped (and even disconnected themselves), the possibility that CSDs whose centroids are close would be non-contiguous is greater. Out of a random sample of 132 rural WAs, 101 (76.5%) were connected, 22 (16.7%) were in two parts and the remaining nine (6.8%) were in more than two parts. Since connectedness is not crucial to the effectiveness of the weighting procedures, the level of connectedness achieved is judged to be satisfactory.

The extent to which WAs respected the boundaries of various census geographic areas was examined separately for (a) CTs in census-tracted areas, (b) CSDs in non-tracted areas. It should be noted that CD (or county) boundaries were always respected.

There were 32 census metropolitan areas (CMAs) or census agglomerations (CAs) in which census tracts were formed in the 1976 Census. In the major cities within these CMAs or CAs, the majority of CTs became WAs. Larger CTs were sometimes divided into two WAs, or in rare cases, three or four WAs. Small CTs were either combined together to form one WA or combined with a neighbouring large CT with the combination being split into two WAs. Some examples of how some municipalities in tracted areas were split into WAs are given in Table 4.

In the smaller municipalities within these CMAs and CAs, the weighting procedures gave priority to municipality boundaries rather than CT boundaries. Therefore, the correspondence between CTs and WAs was less pronounced in these areas.

VI. ÉVALUATION DES PROCÉDURES DE PONDERATION ET DE LA COMPATIBILITÉ DE L'ÉCHANTILLON ET DE LA POPULATION

6.1 Formation des régions de pondération

La première étape de la pondération consistait en la création de régions de pondération (voir le paragraphe 3.4). On a réparti les SD échantillonnés en 4,798 RP ayant une population moyenne de 4,516 habitants (sans compter les personnes résidant dans les logements collectifs). Parmi ces 4,798 RP, 4,672 (97.5 %) avaient une population de 3,000 - 7,000 habitants, tandis que 98 des 126 autres ont été rectifiées manuellement afin de les faire entrer dans ces limites. Les 28 RP restantes avaient toutes une population de 2,500 - 3,000 habitants mais ont été laissées telles quelles pour ne pas modifier davantage les limites des secteurs de recensement (SR) et des subdivisions de recensement (SDR).

Seule la contiguïté des RP des régions rurales a été analysée, étant donné qu'aucun problème ne se posait dans les régions urbaines où les RP correspondaient à des secteurs de recensement compacts. Dans les régions rurales, les SDR peuvent être de grande taille et de forme irrégulière, ou même morcelées, si bien que les SDR dont les centroïdes sont rapprochés risquent fort de n'être pas contiguës. Dans un échantillon aléatoire de 132 RP rurales, 101 (76.5 %) étaient contiguës, 22 (16.7 %) étaient partagées en deux et les neuf autres (6.8 %) comportaient plus de deux parties. La contiguïté n'étant pas un critère essentiel d'efficacité de la pondération, les résultats obtenus sont jugés satisfaisants.

On a étudié dans quelle mesure les RP respectaient les limites des diverses régions géographiques de recensement en fonction a) des SR, pour les régions divisées en secteurs de recensement et b) des SDR, pour les régions ne comportant pas de secteurs de recensement. Il convient de souligner que les limites des DR (ou des comtés) ont toujours été respectées.

Au recensement de 1976, 32 régions métropolitaines de recensement (RMR) ou agglomérations de recensement (AR) ont été divisées en secteurs de recensement. Dans le cas des grandes villes appartenant à ces RMR ou AR, la plupart des SR sont devenus des RP. Les plus gros SR ont parfois été divisés en deux RP, ou même, dans quelques rares cas, en trois ou quatre RP. Les SR de petite taille ont été, soit groupés pour constituer une RP, soit adjoints à un SR voisin plus grand pour être divisés en deux RP. Le tableau 4 illustre comment certaines municipalités situées dans des SR ont été partagées en RP.

Pour ce qui est des municipalités de plus petite taille dans ces RMR et dans ces AR, les procédures de pondération ont donné la priorité aux limites des municipalités plutôt qu'à celles des SR. Par conséquent, la correspondance entre les SR et les RP est moins prononcée dans ces régions.

TABLE 4. Formation of WAs from CTs in Selected Municipalities Within CMAs/CAs
TABLEAU 4. Formation des RP à partir de SR dans les municipalités choisies à l'intérieur des RMR/AR

Municipality — Municipalité	Total number of CTs within municipality (excluding CTs with no pop.) — Nombre total de SR dans la municipalité (excepté les SR sans population)	Number of CTs that — Nombre de SR qui			Other cases — Autres cas	Percentage of CTs with boundaries respected ¹ — Pourcentage de SR dont les limites ont été respectées ¹
		Became one WA — Ont constitué un RP (1)	Were split into two or more WAs — Ont été séparés en deux RP ou plus (2)	Were combined in pairs to form one or more WAs — Ont été regroupés par deux pour constituer une RP ou plus (3)		
Calgary	84	59	13	12	—	85.7
Edmonton	82	61	15	6	—	92.7
Halifax	27	12	1	14	—	48.1
Laval (Montréal)	48	29	7	12	—	75.0
London	49	31	6	12	—	75.5
Moncton	14 ²	7	—	4	3	50.0
Montréal	287	124	15	106	42 ³	48.4
Québec	48 ⁴	13	3	20	12 ⁵	33.3
Saskatoon	24	16	4	4	—	83.3
Scarborough (Toronto) . . .	53	30	21	2	—	96.2
Toronto	139	69	18	38	14 ⁶	62.6
Vancouver	72 ⁷	46	14	10	2 ⁸	83.3
Windsor	43	23	4	16	—	62.8
Winnipeg	115	58	13	40	4 ⁹	61.7
York, East (Toronto)	17	4	5	8	—	52.9
York, North (Toronto) . . .	78	35	35	8	—	89.7

¹ Columns (1) and (2) as a percentage of total number of CTs. — La somme des colonnes (1) et (2) comme pourcentage du nombre total de SR.

² Including three CTs that also contained parts of other municipalities. — Y compris trois SR contenant aussi des parties d'autres municipalités.

³ Twenty-four were combined into six sets of three or more to form one WA per set; 18 were combined into four sets of four or more to form two WAs per set. — Vingt-quatre furent combinés comme suit: six ensembles de trois ou plus pour former une RP par groupe; 18 furent regroupés en quatre groupes de quatre ou plus pour former deux RP par groupe.

⁴ Including two CTs that also contained parts of other municipalities. — Y compris deux SR qui contenaient aussi des parties d'autres municipalités.

⁵ Two sets of three CTs formed two WAs; one set of six CTs formed four WAs. — Deux ensembles de trois SR ont formé deux RP; un ensemble de six SR a formé quatre RP.

⁶ One set of four CTs and one set of six CTs each formed one WA; one set of four CTs formed two WAs. — Un ensemble de quatre SR et un ensemble de six SR ont formé une RP chacun; un ensemble de quatre SR a formé deux RP.

⁷ Excludes Indian Reserves. — Ne comprend pas les réserves indiennes.

⁸ Two CTs combined with another municipality to form one WA. — Deux SR ont été agrégés à une autre municipalité pour former une RP.

⁹ All were combined to form one WA. — Toutes combinées pour former une RP.

Outside the census-tracted areas, the preservation of CSD boundaries depended greatly on the size of CSDs. The majority (approximately 75%) of CSDs in non-tracted areas were grouped into WAs that contained two or more whole CSDs. Less than 10% of CSDs in non-tracted areas were preserved as one or more whole WAs. The remaining CSDs were combined in groups of two or more which were then split into two or more WAs.

6.2 Cross-classification and Collapsing

This section provides a brief summary of the amount of collapsing that took place for the rows and columns of the two weighting matrices (see Appendix II). The amount of collapsing affects the level of agreement achieved between sample estimates and complete counts since the raking ratio estimation procedure only controls the estimates for the rows and columns formed after collapsing. This summary is based on an inspection of about 3,400 WAs.

Since collapsing was carried out separately within each WA, and given the collapsing strategy, many different patterns of collapsing occurred. To summarize, we show in Table 5 the most frequent collapsing that occurred in each dimension of each matrix.

The main observations from Table 5 are:

- (a) Young "never-married" and adult "ever-married" groups for both sexes were almost never collapsed.
- (b) The "Other" mother tongue category was almost always combined with one or both of the two official language mother tongue groups.
- (c) The subcategories of family households with female heads, and of one-person non-family households were frequently collapsed.
- (d) The owned and rented categories were almost always preserved while the type of dwelling categories were sometimes combined.

6.3 Sample Population Consistency

Table 6 summarizes the level of agreement between census totals and the corresponding weighted sample estimates for a selection of categories for Canada. Corresponding tables for the provinces can be found in Appendix IV.

Table 6 shows exact agreement (except for random rounding) in every province for total number of households, total population by sex, and total population by age in the age groups 0 - 14, 15 - 64 and 65 +. Agreement is very close for households by tenure, for single dwellings, for married persons, and for the English and French mother tongue groups in provinces

Dans les régions ne comportant pas de secteurs de recensement, on a plus ou moins respecté les limites des SDR selon la taille de ces dernières. Dans la majorité de ces cas (environ 75 % des SDR), les RP regroupaient deux SDR complètes ou plus. Moins de 10 % des SDR des régions ne comportant pas de secteurs de recensement ont été converties telles quelles en une ou plusieurs RP. Le reste des SDR a été combiné en groupes de deux ou plus pour être divisées en deux RP ou plus.

6.2 Classement recoupé et agrégation

Le présent paragraphe est un bref résumé des agrégations effectuées au niveau des lignes et des colonnes des deux matrices de pondération (voir l'appendice II). Ces agrégations influent sur le degré de concordance des estimations de l'échantillon et des totaux de l'ensemble de la population car la méthode itérative d'estimation des quotients n'a d'effet que sur les estimations des lignes et des colonnes formées après l'agrégation. Ce résumé a été fait à partir de l'étude de quelque 3,400 RP.

L'agrégation ayant été effectuée à l'intérieur de chaque RP séparément, et étant donné la stratégie d'agrégation, on a assisté à différents types d'agrégation. Le tableau 5 reprend les groupes d'agrégation les plus fréquents par rapport à chacune des dimensions des deux matrices.

Les remarques relatives au tableau 5 sont les suivantes:

- a) Il n'y a eu pour ainsi dire aucune agrégation des groupes de jeunes "jamais mariés" et d'adultes "non célibataires" des deux sexes.
- b) La catégorie "Autre" de la langue maternelle a presque toujours été combinée avec l'un ou l'autre des deux groupes de langue maternelle officielle.
- c) Les sous-catégories des ménages familiaux avec un chef féminin et des ménages non familiaux d'une seule personne ont fréquemment été agrégées.
- d) En général, les catégories de propriétaires et locataires n'ont pas été touchées, tandis que les catégories de types de logement ont quelquefois été agrégées.

6.3 Compatibilité de l'échantillon et de la population

Le tableau 6 donne un résumé du degré de concordance entre les totaux du recensement et les estimations pondérées correspondantes de l'échantillon, selon certaines catégories à l'échelle du Canada. On peut retrouver à l'appendice IV les tableaux correspondant au niveau des provinces.

Selon le tableau 6, il y a concordance parfaite (sauf pour l'arrondissement aléatoire) dans chaque province, du nombre total de ménages, de la population totale selon le sexe et de la population totale selon l'âge pour les groupes d'âge 0 - 14 ans, 15 - 64 ans et 65 ans et plus. La concordance des catégories suivantes est presque parfaite: ménages selon le mode d'occupation, logements indivi-

TABLE 5. Most Frequent Collapsing Patterns
(Refer to Appendix II for row and column definitions)

TABLEAU 5. Patterns de regroupement les plus fréquents
(Voir la définition des lignes et des colonnes à l'appendice II)

Rows — Lignes			Columns — Colonnes		
Final row groupings — Groupement final des lignes	Percentage of WAs in which this grouping occurred — Pourcentage de RP dans lesquelles ces groupements se sont produits	Most frequent groupings in remaining WAs ¹ — Groupements les plus fréquents dans les RP restantes ¹	Final column groupings — Groupement final des colonnes	Percentage of WAs in which this grouping occurred — Pourcentage de RP dans lesquelles ces groupements se sont produits	Most frequent groupings in remaining WAs ¹ — Groupements les plus fréquents dans les RP restantes ¹

A. Population and family weight matrix

A. Matrice des coefficients de pondération de la population et des familles

1	99.1	—	1	62.0	or/ou { 1 + 2 + 3 2, 1 + 3
2	99.0	—	2 + 3	57.6	
3	99.2	—	4	73.1	or/ou { 4 + 5 + 6 4, 5, 6
4	99.6	—	5 + 6	59.8	
5	96.7	—	7 + 8 + 9	85.2	or/ou { 10 + 11 + 12 11, 10 + 12
6 + 7 + 8 + 9	64.1	6 + 7, 8 + 9	10	47.2	
10 + 17	92.5	10, 17	11 + 12	44.9	or/ou { 13 + 14 + 15 13, 14, 15
11 + 12	87.5	11 + 12 + 13	13	73.6	
13	87.5	—	14 + 15	59.2	or/ou { 16 + 17 + 18 17, 16 + 18
14	99.9	—	16	65.1	
15	99.2	—	17 + 18	54.3	or/ou { 19 + 20 + 21 20, 19 + 21
16	99.2	—	19	50.3	
18	99.1	—	20 + 21	44.9	or/ou { 22, 23 + 24 23, 22 + 24
19	99.1	—	22 + 23 + 24	46.1	
20	99.3	—	25 + 26 + 27	73.3	25, 26 + 27 28, 29 + 30
21	94.4	with/avec 22 + 26	28 + 29 + 30	81.6	
22	75.9	with/avec 23 + 26			
23 + 24 + 25 + 26	55.0	with/avec 22 or/ou 23 + 24, 25 + 26			
27 + 34	88.5	27, 34			
28 + 29	87.1	28, 29			
30	97.4	—			
31	100.0	—			
32	99.2				
33	99.2				

B. Household and housing weight matrix

B. Matrice des coefficients de pondération des ménages et des logements

1 + 7 + 8	60.4	1, 7 + 8	1	89.5	{ 1 + 2
2	92.6	1 + 2 + 7 + 8	2	89.5	
3	70.3	3 + 9 + 10	3	68.4	{ 3 + 4
4	88.5	3 + 4	4	68.4	
9 + 10	69.8	3 + 9 + 10			
5 + 6 + 11 + 12	59.0	6, 5 + 11 + 12			
13 to/à 18	22.8	13 + 14 + 16 + 17, 15, 18 or/ou 13 + 16, 14 + 17, 15 + 18			
19 + 20	78.2	19, 20			

¹ Where the percentage of WAs in which the most frequent grouping did not occur exceeds 5%, this column indicates what happened in most of the other WAs (x, y means rows x and y remained separate, x + y means they are combined). — Lorsque le pourcentage de RP dans lesquelles les groupements les plus fréquents n'ont pas eu lieu dépasse 5 %, cette colonne indique ce qui s'est produit dans la plupart des autres RP (x, y veut dire que les lignes x et y sont demeurées intactes et x + y veut dire qu'elles ont été groupées).

TABLE 6. Differences Between Census Totals and Corresponding Sample Estimates for Selected Categories, for Canada

TABLEAU 6. Écarts entre les totaux du recensement et les estimations échantillonnelles correspondantes de catégories choisies au niveau du Canada

Category Catégorie	Census total Total du recensement	Estimate from census sample Estimation à partir de l'échantillon de recensement	Percentage difference ¹ $\frac{(2) - (1)}{(1)} \times 100\%$ Écart de pourcentage ¹ $\frac{(2) - (1)}{(1)} \times 100\%$
	(1)	(2)	
Total:			
Households - Ménages	7,166,095	7,166,095	-
Population	22,992,605	22,992,605	-
Families - Familles	5,727,895	5,734,285	0.11
Households of size 1 - 2 - Ménages d'une ou deux personnes . . .	3,195,480	3,197,195	0.05
Households of size 6 and over - Ménages de 6 personnes et plus	656,150	652,390	- 0.57
Owned households - Ménages propriétaires	4,431,235	4,431,150	-
Households occupying single dwellings - Ménages occupant des logements individuels	3,991,545	3,989,980	- 0.04
Household heads aged 25 - 34 - Chefs de ménage âgés de 25 - 34 ans	1,678,965	1,680,815	0.11
Males - Hommes	11,449,520	11,449,520	-
Females - Femmes	11,543,080	11,543,080	-
Persons aged - Personnes âgées:			
0 - 14	5,896,180	5,896,175	-
30 - 34	1,627,485	1,622,405	- 0.31
65 +	2,002,335	2,002,335	-
Married persons - Personnes mariées	10,973,905	10,967,745	- 0.06
Widowed persons - Veufs, veuves	1,043,560	1,048,135	0.44
Persons with English mother tongue - Personnes dont la langue maternelle est l'anglais	14,122,770	14,115,295	- 0.05
Persons with French mother tongue - Personnes dont la langue maternelle est le français	5,887,205	5,884,775	- 0.04
Two-parent families - Familles biparentales	5,168,565	5,175,215	0.13
Lone-parent families - Familles monoparentales	559,330	559,070	- 0.05
Families with no children - Familles sans enfants	1,726,410	1,729,205	0.16

¹ Differences of ± 5 are neglected since they could arise from random rounding. - Il n'est pas tenu compte des écarts de ± 5 étant donné qu'ils peuvent découler d'un arrondissement aléatoire.

where these are large. Generally, the larger the category the closer the agreement in percentage terms. There is an overestimate of about 6,400 families from the census sample at the Canada level. Categories 18 and 19 show that this difference occurs almost exclusively among two-parent families. The weighting matrix used for population and family weights (see Appendix II) used family status in its columns. Since the iterative weighting procedure ends with the rows of the weighting matrix, agreement for families is not expected to be exact. However, the observed difference of about 6,400 families is larger than expected and requires explanation. It appears that this difference stems from the processing problem mentioned in Chapter V in connection with mother tongue. This "pick-up" problem caused some false relationship to head codes to be created on short form questionnaires. These in turn affected the formation of families leading generally to a reduction in the number of families formed from short form questionnaires. Although the weighting procedures attempt to correct for the resultant imbalance in family counts on short forms and long forms, the number of iterations was apparently not sufficient to remove the discrepancy entirely.

To examine the level of sample population consistency at the CD level, data from the 46 CDs in the Atlantic provinces were examined. Table 7 shows the percentage of CDs that showed sample population differences of (a) less than 1%, (b) 1%, but less than 3%, or (c) 3%, but less than 5%, (d) 5% or more, for a variety of categories.

Table 8 shows the results of a similar examination of sample population consistency at the EA level and is based on data for EAs in the Atlantic provinces. Because of the smaller figures involved at the EA level, a different set of percentage ranges have been used.

The level of agreement at the EA level depends heavily on the size of the category in the EA. The majority of sample population differences in excess of 20% occur in EAs where the particular category is very small. For example, among those EAs that show differences of greater than 20% for total population, almost two-thirds have a population of less than 50 persons. The magnitude of these differences is a reflection of the inherent variability of sample estimates for categories having very few units in the sample.

duels, personnes mariées, langue maternelle française et langue maternelle anglaise, dans les provinces où ces groupes étaient en nombre important. En général, plus la catégorie est importante, plus le pourcentage de concordance est grand. L'échantillon du recensement donne une surestimation d'environ 6,400 familles à l'échelle du Canada. Les catégories 18 et 19 révèlent que cet écart se produit presque exclusivement au niveau des familles biparentales. La matrice des coefficients de pondération utilisée pour la population et les familles (voir l'appendice II) emploie la situation des particuliers au sein des familles pour faire des compressions sur ses colonnes. La dernière itération du processus RREP se terminant avec les lignes de la matrice, il pourrait ne pas y avoir de concordance parfaite dans le cas des familles. Toutefois, l'écart observé d'environ 6,400 familles est supérieur aux prévisions et mérite une explication. Cet écart est dû au problème de dépouillement des données relatives à la langue maternelle mentionné au chapitre V. Ce problème de saisie a induit des erreurs dans la création de codes pour les questionnaires abrégés. Cela a ensuite eu pour effet de réduire le nombre de familles formées à partir des questionnaires abrégés. Bien que les procédures de pondération tendent à corriger l'écart entre le nombre de familles calculé à partir des questionnaires complets et abrégés, le nombre d'itérations s'est révélé insuffisant pour l'éliminer entièrement.

L'étude de la compatibilité de l'échantillon et de la population au niveau des DR a porté sur les 46 DR des provinces de l'Atlantique. Le tableau 7 donne les pourcentages de DR qui présentent des écarts entre l'échantillon et la population de l'ordre de a) moins de 1 %, b) 1 à 3 %, c) 3 à 5 % ou d) 5 % ou plus, pour certaines catégories.

Le tableau 8 présente les résultats d'une étude semblable faite au niveau des SD à partir des données des provinces de l'Atlantique. Les chiffres se rapportant aux SD étant moins grands, on a utilisé sur ce tableau des échelles de pourcentages différentes.

Au niveau des SD, le degré de concordance des résultats dépend surtout de l'importance numérique de la catégorie observée dans le SD. La majorité des écarts de 20 % et plus se produisent dans des SD pour lesquels le nombre de personnes appartenant à cette catégorie est très petit. Ainsi, près des deux tiers des SD qui présentent des écarts supérieurs à 20 % au niveau de la population totale ont une population inférieure à 50 personnes. L'amplitude de ces écarts est le reflet de la variabilité propre aux estimations échantillonales de catégories peu importantes numériquement parlant dans l'échantillon.

TABLE 7. CDs by Size of Sample Population Difference for Various Categories for the 46 CDs in the Atlantic Provinces
(Percentages are Shown in Brackets)

TABLEAU 7. DR selon l'écart entre l'échantillon et la population pour diverses catégories des 46 DR des provinces de l'Atlantique
(Les pourcentages sont indiqués entre parenthèses)

Category — Catégorie	Sample population differences — Écarts entre l'échantillon et la population			
	Less than 1% — Moins de 1 %	1% but less than 3% — De 1 % à 3 %	3% but less than 5% — De 3 % à 5 %	More than 5% — Plus de 5 %
Males aged 5-9 — Garçons âgés de 5-9 ans	46 (100.0%)	—	—	—
Females aged 20-24 — Femmes âgées de 20-24 ans	27 (58.7%)	15 (32.6%)	3 (6.5%)	1 (2.2%)
Persons aged 50-54 — Personnes âgées de 50-54 ans	12 (26.1%)	19 (41.3%)	11 (23.9%)	4 (8.7%)
Persons with English mother tongue — Personnes dont la langue maternelle est l'anglais	45 (97.8%)	1 (2.2%)	—	—
Persons with French mother tongue — Personnes dont la langue maternelle est le français	17 (37.0%)	11 (23.9%)	4 (8.7%)	14 (30.4%)
Widowed persons — Veufs, veuves	14 (30.4%)	18 (39.1%)	12 (26.1%)	2 (4.3%)
Divorced persons — Divorcés, divorcées	1 (2.2%)	7 (15.2%)	7 (15.2%)	31 (67.4%)
Single females — Femmes célibataires	46 (100.0%)	—	—	—
All families — Ensemble des familles	46 (100.0%)	—	—	—
Widowed family heads — Chefs de famille veufs	5 (10.9%)	15 (32.6%)	9 (19.6%)	17 (37.0%)
Family heads aged less than 25 — Chefs de famille âgés de moins de 25 ans	13 (28.3%)	16 (34.8%)	12 (26.1%)	5 (10.9%)
Households of size 1-2 — Ménages d'une ou deux personnes	33 (71.7%)	12 (26.1%)	1 (2.2%)	—
Households of size 6 and over — Ménages de 6 personnes et plus	8 (17.4%)	21 (45.7%)	9 (19.6%)	8 (17.4%)
Households occupying single dwellings — Ménages occupant des logements individuels	46 (100.0%)	—	—	—

TABLE 8. Percentage Distribution of EAs by Size of Sample Population Difference for Various Categories for EAs in the Atlantic Provinces¹

TABLEAU 8. Répartition en pourcentage des SD selon l'écart entre l'échantillon et la population pour diverses catégories des SD des provinces de l'Atlantique¹

Category — Catégorie	Sample population differences — Écarts entre l'échantillon et la population			
	Less than 5% — Moins de 5 %	5% but less than 10% — De 5 % à 10 %	10% but less than 20% — De 10 % à 20 %	More than 20% — Plus de 20 %
Total population — Population totale	60.5	24.7	10.4	4.4
Persons aged 15-24 — Personnes âgées de 15-24 ans	27.4	22.2	27.6	22.8
Single persons — Personnes seules	38.6	27.3	23.0	11.1
Persons with English mother tongue — Personnes dont la langue maternelle est l'anglais	52.6	24.3	13.0	10.2
All families — Ensemble des familles	65.2	23.6	7.8	3.5
Families with heads aged 35-64 — Familles ayant un chef âgé de 35-64 ans	34.0	26.6	25.4	14.0
Households of size 1-2 — Ménages d'une ou deux personnes	13.9	15.8	24.9	45.3
Owner-occupied dwellings — Logements occupés par le propriétaire	69.6	20.7	6.0	3.7
Households occupying single dwellings — Ménages occupant des logements individuels	41.6	27.1	19.5	11.8

¹ Based on all EAs in the Atlantic provinces except those in which sampling was not used and those having no persons, families, households, or dwellings in the given category. — D'après les données de tous les SD des provinces de l'Atlantique sauf ceux qui n'ont pas fait l'objet d'un échantillonnage et ceux où il n'y avait pas de personnes, familles, ménages ou logements des catégories données.

VII. THE CONTRIBUTION OF SAMPLING TO TOTAL VARIANCE

7.1 Total Variance

The reliability of an estimate made from a survey or a census can be summarized by its total variance which measures the variability of the estimate about its average value in hypothetical repetitions of the survey process. An estimate that would always take on very similar values if the survey (including the sampling process) were repeated under identical conditions would have a low total variance, whereas an estimate that varied greatly in repetitions would show a high total variance. Variability due to sampling results from the random choice of units to be in the sample. Generally, the larger the sample the less the variability due to sampling. Variability due to non-sampling causes can arise from random errors occurring at the response stage or during processing. Some questions on the questionnaire may give rise to response error either because of the concepts they address or as a result of their presentation. Some random errors will be introduced into processing operation (e.g., coding) even though Quality Control procedures are used to ensure that such errors are contained within tolerable bounds. Both of these sources of non-sampling variability contribute to the total variance of estimates whether they are based on a complete enumeration or on a sample. It should be noted that total variance does not include the effect of any bias in the estimates. Bias is defined as the difference between the average value of the estimate in hypothetical repetitions of the survey process and the true value being estimated. It represents the components of error that do not tend to cancel out over such repetitions. Bias may arise from undercoverage, from response errors, or from processing errors.

The purpose of this chapter is to examine the effect of sampling on the total variance of those census figures that are based on sample data. Unfortunately, it is impossible to separate completely the contributions of sampling and non-sampling errors to total variance unless one actually conducts an identical and independent repetition of the census process (at least for a sample of areas). Although this could not be done in conjunction with the 1976 Census, it was possible to produce a separate estimate of the contribution to total variance from a significant component of non-sampling error, correlated response variance. The non-sampling error component can be split into two components: (a) simple response variance which measures

VII. INCIDENCE DE L'ÉCHANTILLON SUR LA VARIANCE TOTALE

7.1 Variance totale

La fiabilité des estimations faites à partir d'une enquête ou d'un recensement est évaluée à partir de leur variance totale qui mesure la variabilité de l'estimation par rapport à sa valeur moyenne obtenue pour des répétitions hypothétiques du processus d'enquête. Une estimation qui garderait toujours à peu près la même valeur si l'enquête (et le processus de sondage) était répétée dans des conditions identiques aurait une faible variance totale, tandis qu'une estimation qui varierait beaucoup aurait une variance totale élevée. La variabilité attribuable à l'échantillonnage vient de ce que les unités qui composent l'échantillon sont choisies au hasard. En général, plus l'échantillon est grand, plus la variabilité échantillonnale est faible. La variabilité qui n'est pas due à l'échantillonnage peut être causée par des erreurs aléatoires commises au niveau de la réponse ou à l'occasion du dépouillement. Certaines questions du questionnaire peuvent donner lieu à des réponses erronées, soit en raison des concepts en cause, soit en raison de leur formulation. Un certain nombre d'erreurs aléatoires seront produites au moment du dépouillement (par exemple au codage) en dépit de l'application de procédures de contrôle de la qualité visant à maintenir leur nombre dans des limites tolérables. Ces deux sources de variabilité non due à l'échantillonnage contribuent à la variance totale des estimations qu'elles soient faites à partir d'un échantillon aussi bien que d'un recensement intégral. Il convient de rappeler que la variance totale ne rend pas compte de l'effet du biais sur les estimations. Par définition, le biais est l'écart entre la valeur moyenne de l'estimation dans l'hypothèse d'une répétition de la procédure d'enquête et la valeur réelle correspondante. Il représente les composantes d'erreur qui ne s'annulent pas au long des répétitions. Le biais peut être le résultat d'un dénombrement incomplet, d'erreurs de réponse ou d'erreurs de traitement.

Le présent chapitre a pour objet d'étudier l'incidence de l'échantillonnage sur la variance totale des estimations du recensement obtenues à partir de données échantillonnales. Malheureusement, il est impossible de déterminer avec précision les contributions réciproques des erreurs dues à l'échantillonnage et des erreurs non dues à l'échantillonnage à moins de procéder à une répétition indépendante du recensement dans des conditions identiques (du moins pour un échantillon de régions). Bien que cela n'ait pu être fait dans le cas du recensement de 1976, il a été possible de produire une estimation de la contribution d'une composante significative de l'erreur non due à l'échantillonnage à la variance totale c'est-à-dire de la variance de réponse corrélée. La

the effect of response errors occurring independently in different households; (b) correlated response variance which measures the effect of response errors that are correlated for different households (e.g., as a result of some common error-causing agent affecting different households in a similar way). Correlated response variance, which may be thought of as an interviewer effect, has been shown in earlier studies to be a major contributor to total variance in interview surveys. Since 1971, the census has been conducted largely by self-enumeration rather than interviews and therefore the interviewer effect is much smaller than in previous canvasser censuses. Nevertheless, the CR can still influence the data during the editing and follow-up stage.

7.2 The Contribution of Sampling

The figures in Table 9 indicate the estimated breakdown of total variance into two components: (a) sampling variance and simple response variance, (b) correlated response variance for a selection of categories.¹¹

These figures are themselves estimates and are presented only to give a general indication of the relative size of the components of total variance. For most of the categories considered, sampling (and simple response) variance account for between 75% and 90% of total variance. (There were also some smaller categories, not shown in Table 9, for which the estimated correlated response variance was negative.) The split of total variance varies by category as expected. There is a tendency for sampling variance to account for a higher percentage of total variance for those categories that are clustered by household (e.g., migration categories). Conversely, the highest contributions by correlated response variance occur for the labour force categories where the subtlety of the concepts and definitions used may leave a greater potential for response error.

It is of some interest to note that the model for total variance which was used in predicting the reliability of both 1971 and 1976 Census sample estimates implied a 73% and 27% split of total variance between sampling variance and response variance. Given that the split in Table 9 has included simple response variance with sampling variance, this model still appears reasonable although it may be slightly over-stating the contribution of response variance.

Estimates of the total variance of census figures have appeared in the individual bulletins of the Volume Series for the benefit of users.

composante de l'erreur non due à l'échantillonnage peut être séparée en deux parties: a) la variance de réponse simple qui mesure les effets des erreurs de réponse produites indépendamment dans différents ménages, et b) la variance de réponse corrélée qui mesure les effets des erreurs de réponse de différents ménages, qui ont une corrélation entre elles (par exemple, celles qui sont dues à une même cause dont les effets se sont répercutés de la même façon sur divers ménages). Des études antérieures ont révélé que la variance de réponse corrélée, qui peut être assimilée aux effets de l'interviewer, influe beaucoup sur la variance totale dans le cas des enquêtes par interview. Depuis 1971, le recensement a utilisé l'auto-dénombrement de préférence aux interviews, le rôle de l'interviewer a donc moins d'effet sur les résultats que dans les premiers recensements par représentant. Toutefois, l'influence du RR peut toujours se faire sentir au moment du contrôle et du suivi.

7.2 Incidence de l'échantillonnage

Le tableau 9 donne une estimation de la répartition de la variance totale entre deux éléments: a) la variance échantillonnale et la variance de réponse simple, d'une part, et b) la variance de réponse corrélée, d'autre part, par rapport à diverses catégories¹¹.

Ces résultats sont eux-mêmes des estimés et ne visent qu'à donner une indication générale de l'importance relative des composantes de la variance totale. Pour la plupart des catégories visées, la variance échantillonnale (et la variance de réponse simple) représente de 75 % à 90 % de la variance totale. (Pour certaines petites catégories qui n'apparaissent pas au tableau 9, les estimations se rapportant à la variance de réponse corrélée étaient négatives.) La répartition de la variance totale varie selon les catégories, ainsi qu'il avait été prévu. La variance échantillonnale tend à représenter un pourcentage plus élevé de la variance totale pour les variables homogènes à l'intérieur des ménages (par exemple, les catégories se rapportant à la migration). Réciproquement, ce sont les catégories de population active qui contribuent le plus à accentuer la variance de réponse corrélée, la complexité de ces concepts et de leurs définitions offrant plus de possibilités d'erreur de réponse.

Il est intéressant de souligner que le modèle de variance totale utilisé pour établir les prévisions concernant la fiabilité des estimations faites à partir d'échantillons pour les recensements de 1971 et 1976 prévoyait une répartition de la variance totale entre la variance échantillonnale et la variance de réponse de l'ordre de 73 % et 27 %. Étant donné que le tableau 9 regroupe la variance de réponse simple et la variance échantillonnale, le modèle reste tout à fait raisonnable bien qu'il surestime légèrement la contribution de la variance de réponse.

Les estimations de la variance totale des données du recensement ont été publiées dans les bulletins de la Série de volumes à l'intention des utilisateurs. Elles sont regroupées dans l'introduction de chaque série de volumes.

¹¹ For a reference to how these figures were obtained, see footnote 6.

¹¹ Pour les références du document expliquant de quelle façon on a obtenu ces résultats, se reporter à la note 6.

TABLE 9. Estimated Breakdowns of Total Variance into (a) Sampling and Simple Response Variance and (b) Correlated Response Variance, for Selected Categories

TABLEAU 9. Estimation de la répartition de la variance totale en fonction a) de la variance échantillonnale et de la variance de réponse simple et b) de la variance de réponse corrélée, pour certaines catégories

Category — Catégorie	Size of category (millions) — Taille de la catégorie (en millions)	Coefficient of variation ¹ — Coefficient de variation ¹	Percentage of total variance due to — Pourcentage de la variance dû à la	
			Sampling and simple response variance — Variance échan- tillonnale et à la variance de réponse simple	Correlated response variance — Variance de réponse corrélée
(1)	(2)	(3)	(4)	(5)
Population in the labour force — Population active. . .	10.26	0.045	79.1	20.9
Employed persons — Personnes actives occupées	9.56	0.046	82.5	17.5
Population not in the labour force — Population inactive.	6.83	0.062	74.8	25.2
Non-movers — Personnes n'ayant pas déménagé	10.93	0.070	89.8	10.2
Movers — Personnes ayant déménagé.	10.31	0.066	93.0	7.0
Migrants	5.32	0.106	87.2	12.8
Population with highest grade less than Grade 5 — Po- pulation ayant moins de 5 années de scolarité.	0.86	0.021	77.5	22.5
Population with highest Grades 5-8 — Population ayant de 5-8 années de scolarité.	3.56	0.088	83.0	17.0
Population with highest Grades 11-13 — Population ayant de 11-13 années de scolarité.	5.79	0.057	94.3	5.7

¹ This is the square root of the total variance of the estimate expressed as a percentage of the estimate. — Représente la racine carrée de la variance totale de l'estimation exprimée comme un pourcentage de l'estimé.

7.3 Comparisons With Simple Random Sampling

Under the assumption of 1 in 3 simple random sampling and simple weighting by 3, the sampling variance of a census sample estimate of the number of persons or households falling in a particular category of a tabulation for a certain geographic area can be expressed as a function of the size of the category and the size of the geographic area. This simple random sample approach to obtaining the sampling variance ignores the following aspects of the sample design and weighting procedure:

- (a) If the cell is a count of persons, then the sample is a cluster sample, the household being the cluster. This will tend to increase the sampling variance compared to a simple random sample especially for variables such as migration for which persons in the same household tend to be similar.

7.3 Comparaison avec l'échantillon aléatoire simple

Dans l'hypothèse d'un échantillon aléatoire simple de 1 sur 3 et d'une pondération simple par 3, la variance échantillonnale de l'estimation, faite à partir d'un échantillon du recensement, du nombre de personnes ou de ménages appartenant à une catégorie donnée pour une région géographique particulière peut être exprimée en fonction de la taille de la catégorie et celle de la région géographique. Cette façon d'obtenir la variance échantillonnale à partir d'un échantillon aléatoire simple ne tient pas compte des aspects du plan de sondage et de la procédure de pondération mentionnés ci-après:

- a) Si la cellule correspond à un nombre de personnes, alors l'échantillon est constitué de grappes, chaque grappe correspondant à un ménage. Cela a tendance à accroître la variance échantillonnale par rapport à celle d'un échantillon aléatoire simple, particulièrement dans le cas de variables telles que la migration où les membres d'un même ménage tendent à être semblables.

- (b) The sample is actually stratified by EA. For geographic areas larger than the EA, this tends to reduce the sampling variance compared to simple random sampling.
- (c) The sample is actually selected systematically within EAs which, compared to simple random sampling, tends to reduce the sampling variance for dwelling variables.

By comparing the sampling variances obtained under these simple assumptions with the figures given in Table 9, one can obtain a measure of the net effect of the above three factors on sampling variance. This comparison does not include the effect of the weighting procedures in reducing the sampling variance, since the sampling variance estimates also assumed simple weighting by three. However, independent studies of the effect of the weighting procedures showed reductions of anywhere from 5% to 25% in the standard error for labour force, education, and migration categories compared to simply weighting by three. Table 10 presents the results of this comparison for a selection of categories. These results are presented in terms of standard errors (i.e., the square roots of the variance figures) so that they are in the same units as the estimates to which they relate.

Ratios of the kind shown in Column (5) of Table 10 are known as design effects (abbreviated to defts) since they show the effect of the sample design on the standard error of sample estimates. As mentioned above, the figures in Column (3) exclude the effect of the raking ratio estimation procedure and include some effects of non-sampling error. Therefore, the absolute values of the defts are of less significance than their relative sizes for different categories. Among person categories, Table 10 shows the largest deft values for the migration characteristics. This reflects the effect of the clustering by household at sample selection. Since the migration status is often the same for all persons in the same household, a cluster sample results in a higher standard error than a simple random sample of the same size. For the labour force and education categories, however, the clustering effect is much smaller. When parents in families are considered, the defts decrease for all categories but are still largest for migration categories. When only household heads are considered, there is no effect of clustering and all characteristics exhibit defts of approximately the same magnitude.

- b) L'échantillon est stratifié en fonction des SD. Dans le cas d'une région géographique plus grande que le SD, la variance échantillonnale tend à être réduite par rapport à celle d'un échantillon aléatoire simple.
- c) L'échantillon est pris systématiquement à l'intérieur des SD ce qui tend à réduire la variance échantillonnale des variables se rapportant au logement, par rapport à l'échantillon aléatoire simple.

On obtient la mesure de l'effet net de ces trois facteurs sur la variance échantillonnale en comparant les variances échantillonnales obtenues à partir de ces hypothèses simples avec celles fournies par le tableau 9. Une telle comparaison n'inclut pas les effets de la procédure de pondération qui contribue à réduire la variance échantillonnale étant donné que les estimations de la variance échantillonnale sont également faites en fonction d'une pondération simple de trois. Toutefois, d'après des études indépendantes de l'incidence des procédures de pondération, il y aurait une réduction pouvant aller de 5 % à 25 % de l'erreur type pour les catégories de population active, scolarité et migration, par rapport à la pondération simple de trois. Le tableau 10 donne les résultats de cette comparaison pour certaines catégories. Ces résultats sont donnés en termes d'erreur type (c'est-à-dire la racine carrée de la variance) de telle sorte qu'ils sont donnés dans les mêmes unités que les estimations auxquelles ils correspondent.

Les quotients de la colonne (5) du tableau 10 correspondent à l'incidence du plan de sondage sur l'erreur type des estimés échantillonnals. Ils sont désignés par l'abréviation anglaise "defts". Ainsi que nous l'avons mentionné plus haut, les chiffres de la colonne (3) ne tiennent pas compte de l'incidence de la procédure d'estimation mais toutefois rendent compte dans une certaine mesure de l'incidence des erreurs qui ne sont pas dues à l'échantillonnage. Par conséquent, la valeur absolue de defts est moins significative que son importance relative dans chaque catégorie. En ce qui concerne les catégories se rapportant aux personnes, les plus grandes valeurs de defts du tableau 10 correspondent aux caractéristiques de la migration. Cela est dû au fait que l'échantillon est constitué de grappes correspondant aux ménages. Le statut migratoire étant généralement le même pour tous les membres d'un même ménage, l'erreur type est plus grande dans un échantillon par grappes que dans un échantillon simple de même taille. Cependant, l'effet de la répartition en grappes est moindre dans le cas des catégories de population active et de scolarité. Lorsqu'il s'agit des parents d'une famille, les valeurs de defts diminuent pour toutes les catégories, mais restent élevées pour les catégories se rapportant à la migration. Lorsqu'on considère uniquement les chefs de ménage, la répartition en grappes est sans effet et les valeurs de defts sont sensiblement du même ordre de grandeur pour toutes les catégories.

TABLE 10. Comparison of Estimated Sampling Error for a Selection of Census Categories With the Sampling Error Assuming Simple Random Sampling

TABLEAU 10. Comparaison de l'erreur échantillonnale estimée pour certaines catégories du recensement avec l'erreur échantillonnale dans l'hypothèse d'un échantillonnage aléatoire simple

Category Catégorie	Size of category (millions) Taille de la catégorie (en millions)	Estimated standard error due to sampling Estimé de l'erreur type due à l'échantillonnage	Standard error under simple random sampling Erreur type pour un échantillonnage aléatoire simple	Ratio of Column (3) to Column (4) Rapport entre la colonne (3) et la colonne (4)
(1)	(2)	(3)	(4)	(5)
A. Person categories Catégories reliées aux personnes				
Population:				
With highest grade less than Grade 5 – Ayant moins de 5 années de scolarité	0.86	1,907	1,287	1.48
With highest Grades 5-8 – Ayant de 5-8 années de scolarité	3.56	2,837	2,454	1.16
With highest Grades 9-10 – Ayant de 9-10 années de scolarité	3.33	2,820	2,387	1.18
With highest Grades 11-13 – Ayant de 11-13 années de scolarité	6.61	3,645	2,943	1.24
Employed persons – Personnes actives occupées . . .	9.56	3,977	3,342	1.19
Population not in the labour force – Population inactive	6.83	3,551	3,100	1.15
Non-movers – Personnes n'ayant pas déménagé . . .	10.93	7,376	3,388	2.18
Non-migrants	4.98	5,021	2,794	1.80
Migrants	5.32	5,092	2,860	1.78
Migrants from same province – Migrants d'une même province	3.49	4,349	2,434	1.79
B. Categories for parents in families Catégories reliées aux parents dans les familles				
Highest Grades 9-10 – Ayant de 9-10 années de scolarité	2.01	2,162	1,811	1.19
Highest Grades 12-13 – Ayant de 12-13 années de scolarité	1.50	2,217	1,606	1.38
In the labour force – Actifs	6.84	3,129	2,256	1.39
Employed persons – Personnes actives occupées . . .	6.51	3,083	2,290	1.35
Not in the labour force – Inactifs	4.06	2,720	2,258	1.20
Non-movers – Personnes n'ayant pas déménagé . . .	5.38	3,626	2,333	1.55
Non-migrants	2.92	2,881	2,068	1.39
Migrants	2.52	3,029	1,933	1.57
C. Categories for household heads Catégories reliées aux chefs de ménage				
Highest grade less than Grade 9 – Ayant moins de 9 années de scolarité	2.20	1,741	1,747	1.00
Highest Grades 9-10 – Ayant de 9-10 années de scolarité	1.20	1,527	1,411	1.08
Highest Grades 11-13 – Ayant de 11-13 années de scolarité	1.33	1,728	1,470	1.18
In the labour force – Actifs	5.36	1,618	1,644	1.02
Not in the labour force – Inactifs	5.14	1,737	1,703	1.02
Employed persons – Personnes actives occupées . . .	1.81	1,665	1,644	1.02
Movers – Personnes ayant déménagé	3.70	1,865	1,892	0.99
Non-migrants	1.84	1,649	1,644	1.00
Migrants from same province – Migrants d'une même province	1.27	1,418	1,446	0.98

VIII. CONCLUSION

Sampling is now an accepted and integral part of census-taking. Its use can lead to substantial reductions in costs and respondent burden associated with a census, or alternatively, can allow the scope of a census to be broadened at the same cost. The price paid for these advantages is an increase in the standard error of census figures that are based on the sample. The effect of sampling is most important for small census figures, whether they are counts for rare categories at the national or provincial level or counts for categories in small geographic areas. It should be noted that response errors and processing errors also contribute to the standard error of census figures and it is the same small census figures that are particularly susceptible to the effects of these non-sampling errors. Therefore, even with a 100% census, many small figures would be of limited reliability. As a general rule of thumb for the 1976 Census, figures of size 50 or less that are based on sample data are of very low reliability, while figures up to size 500 tend to have standard errors in excess of 10% of their size.

The implementation of sampling procedures in the field requires careful preparation and control to ensure that no selection bias is introduced into the sample by the CRs. Despite some apparent deviations from procedures in the 1976 Census, no evidence of selection bias in the sample was found. The procedures for weighting sample data up to the population level generally achieved the levels of sample population consistency anticipated. One exception was the count of two-parent families where a set of circumstances combined to produce a deviation of some 6,000 families (0.13%) at the national level.

VIII. CONCLUSION

L'échantillonnage fait maintenant partie intégrante du processus de recensement. Son utilisation permet de faire des économies substantielles et de réduire de façon appréciable le fardeau du répondant, ou encore, d'élargir la portée d'un recensement sans frais supplémentaires. Le prix de ces avantages est un accroissement de l'erreur type dans le cas des estimés obtenus à partir de l'échantillon. Cet effet se fait particulièrement sentir sur les estimés de faible valeur du recensement, qu'il s'agisse des chiffres se rapportant à des catégories peu usuelles au niveau national ou provincial, ou de ceux correspondant à des régions géographiques de petite taille. Il convient de souligner que les erreurs de réponse et les erreurs de dépouillement contribuent également à l'erreur type des estimés du recensement et, là encore, ce sont les estimés de faible valeur qui sont particulièrement sensibles aux effets de ces erreurs non dues à l'échantillonnage. De toute façon, elles n'auraient qu'une fiabilité relative lors d'un recensement. Pour ce qui est du recensement de 1976, les estimés inférieurs ou égaux à 50 établis à partir des données de l'échantillon sont généralement très peu fiables, tandis que les estimés entre 50 et 500 tendent à avoir des erreurs type de l'ordre de 10 % fois leur valeur.

La mise en application des procédures d'échantillonnage sur le terrain nécessite une préparation sérieuse et un contrôle rigoureux afin d'assurer qu'aucun biais de sélection n'est introduit dans l'échantillon par les RR. Bien que l'on ait relevé des déviations de procédures au recensement de 1976, il n'y a eu aucune preuve d'un tel biais dans l'échantillon. Les procédures de pondération appliquées pour aligner les données de l'échantillon sur celles de l'ensemble de la population ont généralement permis d'atteindre le degré de compatibilité prévu. Il y a eu une exception, toutefois, dans le cas des familles biparentales où un concours de circonstances a occasionné un écart d'environ 6,000 familles (0.13 %) au niveau national.

APPENDIX I

Determining the Random Start

Exhibit 1.1 is a reproduction of pages 2 and 3 of the Visitation Record and shows the shading used to identify households in the sample.

Exhibit 1.2 is a reproduction of the page from the Census Commissioner's Manual which gives instructions for determining the random start in each EA.

APPENDICE I

Détermination aléatoire des points de départ

La pièce 1.1 est une reproduction des pages 2 et 3 du Registre des visites et montre comment les ménages dans l'échantillon ont été identifiés au moyen de cases ombrées.

La pièce 1.2 est une reproduction de la page du Manuel du commissaire au recensement contenant les directives concernant la détermination aléatoire des points de départ dans chaque SD.

SECTION I

- 2 -

PRIVATE DWELLINGS

SINGLE HOMES, APARTMENTS,									
UNINCORPORATED PLACE OR BLOCK NUMBER (if applicable)	HOUSEHOLD NUMBER	EXACT LOCATION OF THIS DWELLING			NAME OF THE HEAD OF THIS HOUSEHOLD	DATE OF FIRST VISIT TO THIS HOUSEHOLD	NUMBER OF PERMANENT (Usual) RESIDENTS	LANGUAGE PREFERENCE OF HOUSEHOLD	
		(1) For cities, towns and villages, enter street and civic number. (Give apt. no. if any.) (2) For other areas, give: (a) Lot, concession and rural route number in Quebec and Ontario. (b) Sec., twp., range and meridian in the Prairies. (c) Parish, twp., etc., in other provinces.			Print surname, given name and initials of Head of household.			7 (✓)	
1	2	3			4	5	6	7 (✓)	
		Civic No.	Street	Apt. No.		Date		Eng.	Fr.
	001		.						
	002		.						
	003		.						
	004		.						
	005		.						
	006		.						
	007		.						
	008		.						
	009		.						
	010		.						
	011		.						
	012		.						
	013		.						
	014		.						
	015		.						
	016		.						
	017		.						
	018		.						
	019		.						
	020		.						
	021		.						
	022		.						
	023		.						
	024		.						
	025		.						
	026		.						
	027		.						
	028		.						
	029		.						
	030		.						
TOTAL DWELL- INGS		Section	Township	Range	Meridian				
		Prairie Provinces							
TOTAL PERMANENT (USUAL) RESIDENTS →									

- OCCUPIED

- 3 -

SECTION I

FLATS, ETC.																
HOUSEHOLD NUMBER	DWELLING ON ONE ACRE OR MORE		\$50 OR MORE OF AGRICULTURAL SALES IN THE PAST 12 MONTHS		AGRICULTURE QUESTIONNAIRE NUMBER	MAIL-BACK ONLY			NUMBER OF TEMPORARY RESIDENTS	COMPLETED HOUSEHOLDS			COMPLETED AGRICULTURAL HOLDINGS		REMARKS	
	Yes	No	Yes	No		12 QUESTIONNAIRE RECEIVED BY MAIL Date	13 RESOLVED BY TEL. FOLLOW-UP Date	14 RESOLVED BY FIELD FOLLOW-UP Date		15	16 (✓)			17 (✓)		
8	9 (✓)		10 (✓)		11	12	13	14	15	16 (✓)			17 (✓)		18	
	Yes	No	Yes	No		Date	Date	Date		2A	2B	4A	6	4A		
001																
002																
003																
004																
005																
006																
007																
008																
009																
010																
011																
012																
013																
014																
015																
016																
017																
018																
019																
020																
021																
022																
023																
024																
025																
026																
027																
028																
029																
030																
TOTAL AGRICULTURAL HOLDINGS →						TOTAL TEMPORARY RESIDENTS →										
										TOTAL COMPLETED HOUSEHOLDS			TOTAL COMPLETED AGRICULTURAL HOLDINGS			

SECTION I

- 2 -

LOGEMENTS PRIVÉS

MAISONS INDIVIDUELLES, APPARTEMENTS,									
LOCALITÉ NON CONSTITUÉE OU ÎLOT N° (s'il y a lieu)	MÉNAGE N°	SITUATION EXACTE DU LOGEMENT			NOM DU CHEF DE CE MÉNAGE	DATE DE LA PREMIÈRE VISITE À CE MÉNAGE	NOMBRE DE RÉSIDENTS PERMANENTS (Habituels)	LANGUE DE PRÉFÉRENCE DU MÉNAGE	
1	2	3			4	5	6	7 (✓)	
		N° de voirie	Rue	App. n°		Date		Ang.	Fran.
	001		.						
	002		.						
	003		.						
	004		.						
	005		.						
	006		.						
	007		.						
	008		.						
	009		.						
	010		.						
	011		.						
	012		.						
	013		.						
	014		.						
	015		.						
	016		.						
	017		.						
	018		.						
	019		.						
	020		.						
	021		.						
	022		.						
	023		.						
	024		.						
	025		.						
	026		.						
	027		.						
	028		.						
	029		.						
	030		.						
LOGE- MENTS, TOTAL		Section	Township	Rang	Méridien	RÉSIDENTS PERMANENTS (HABITUELS), TOTAL →			
		Provinces des Prairies							

- OCCUPÉS

- 3 -

SECTION I

LOGEMENTS DE PLAIN-PIED, ETC.																
MÉNAGE N°	LOGEMENT SUR UNE ACRE OU PLUS		VENTES AU COURS DES 12 DERNIERS MOIS DE PRODUITS AGRICOLES D'UNE VALEUR DE \$50 OU PLUS		QUESTIONNAIRE AGRICOLE N°	RETOUR PAR LA POSTE SEULEMENT			NOMBRE DE RESIDENTS TEMPORAIRES	MÉNAGES - TRAVAIL TERMINÉ			EXPLOITATIONS AGRICOLES - TRAVAIL TERMINÉ	REMARQUES		
	9 (✓)		10 (✓)			11	12	13		14	15	16 (✓)			17 (✓)	
	Oui	Non	Oui	Non		Date	Date	Date		2A	2B	4A	6	4A	18	
001																
002																
003																
004																
005																
006																
007																
008																
009																
010																
011																
012																
013																
014																
015																
016																
017																
018																
019																
020																
021																
022																
023																
024																
025																
026																
027																
028																
029																
030																
EXPLOITATIONS AGRICOLES, TOTAL →					RESIDENTS TEMPORAIRES, TOTAL →											
MÉNAGES - TRAVAIL TERMINÉ, TOTAL										EXPLOITATIONS AGRICOLES - TRAVAIL TERMINÉ, TOTAL						

14. ALLOCATION OF PREDETERMINED STARTING POINTS IN THE VR

In order to satisfy the sampling requirements, the first one or two lines may have to be deleted from Section I of the VR. This should be done with a straight line through all the columns on the page. The deletion will depend on the last two digits of the EA number, e.g., EA 206 will have households 001 and 002 deleted as shown for 06 on the chart below.

14. ATTRIBUTION DES POINTS DE DÉPART PRÉÉTABLIS DANS LE RV

L'échantillonnage exige que, dans certains cas, la première ou les deux premières lignes de la section I du RV soient annulées. Il faut supprimer ces lignes en rayant d'un trait horizontal toutes les colonnes de la page. Ce sont les deux derniers chiffres du numéro du SD qui déterminent quelles lignes sont annulées, par exemple, pour le numéro 206, on enlèvera les numéros 001 et 002 ainsi qu'il est indiqué pour 06 sur le graphique ci-dessous.

Last two digits of EA number — Deux derniers chiffres du numéro du SD	Delete household number — Numéros de ménage à biffer	Last two digits of EA number — Deux derniers chiffres du numéro du SD	Delete household number — Numéros de ménage à biffer
01	—	51	001 and/et 002
02	001	52	—
03	001 and/et 002	53	001
04	—	54	001 and/et 002
05	001	55	—
06	001 and/et 002	56	001
07	—	57	001 and/et 002
08	001	58	—
09	001 and/et 002	59	001
10	—	60	001 and/et 002
11	001	61	—
12	001 and/et 002	62	001
13	—	63	001 and/et 002
14	001	64	—
15	001 and/et 002	65	001
16	—	66	001 and/et 002
17	001	67	—
18	001 and/et 002	68	001
19	—	69	001 and/et 002
20	001	70	—
21	001 and/et 002	71	001
22	—	72	001 and/et 002
23	001	73	—
24	001 and/et 002	74	001
25	—	75	001 and/et 002
26	001	76	—
27	001 and/et 002	77	001
28	—	78	001 and/et 002
29	001	79	—
30	001 and/et 002	80	001
31	—	81	001 and/et 002
32	001	82	—
33	001 and/et 002	83	001
34	—	84	001 and/et 002
35	001	85	—

During drop-off training, you must indicate to each Census Representative which lines, if any, of the VR are to be deleted. You must check each VR to be certain that the correct household numbers have been deleted.

Au cours de la formation pour la livraison, vous devez indiquer à chaque représentant du recensement quelles lignes du RV il doit supprimer (le cas échéant). Vous devez vérifier chaque RV afin de vous assurer que le représentant a biffé les bons numéros de ménage.

APPENDIX – II – APPENDICE

TABLE PF1. Cross-classification Table for Population and Family Weight
TABLEAU PF1. Classement recoupé des coefficients de pondération de la population et des familles

				Family members Personnes faisant partie d'une famille																		Non-family members Personnes hors famille											
				Head in a husband-wife family Chef d'une famille époux-épouse					Head in a one-parent family		Spouse in a husband-wife family Épouse d'une famille époux-épouse					Children in the families Enfants faisant partie d'une famille					Head of household Chef du ménage		Related member of household Parent-membre d'un même ménage		Unrelated member of household Personnes non apparentées dans un ménage								
				Without children Sans enfants		With children Ayant des enfants			Chef d'une famille mono-parentale		11		22			0 - 14 years Âgés de 0 - 14 ans		15 + years Âgés de 15 ans et plus															
Sex Sexe	Marital status État matrimonial	Age Âge	No. No	Mother tongue – Langue maternelle																													
				E. A.	F.	Other Autre	E. A.	F.	Other Autre	E. A.	F.	Other Autre	E. A.	F.	Other Autre	E. A.	F.	Other Autre	E. A.	F.	Other Autre	E. A.	F.	Other Autre	E. A.	F.	Other Autre	E. A.	F.	Other Autre			
				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Male Hommes	Never married Célibataires	0 - 4	1																														
		5 - 9	2																														
		10 - 14	3																														
		15 - 19	4																														
		20 - 24	5																														
		25 - 34	6																														
		35 - 44	7																														
		45 - 54	8																														
		55 - 64	9																														
		65 +	10																														
	Ever married Non célibataires	0 - 19	11																														
		20 - 24	12																														
		25 - 34	13																														
		35 - 44	14																														
		45 - 54	15																														
		55 - 64	16																														
		65 +	17																														
Female Femmes	Never married Célibataires	0 - 4	18																														
		5 - 9	19																														
		10 - 14	20																														
		15 - 19	21																														
		20 - 24	22																														
		25 - 34	23																														
		35 - 44	24																														
		45 - 54	25																														
		55 - 64	26																														
		65 +	27																														
	Ever married Non célibataires	0 - 19	28																														
		20 - 24	29																														
		25 - 34	30																														
		35 - 44	31																														
		45 - 54	32																														
		55 - 64	33																														
		65 +	34																														

¹ Some children less than six years. – Ayant des enfants âgés de moins de six ans.

² No children less than six years. – Aucun enfant âgé de moins de six ans.

APPENDIX II – Concluded – APPENDICE II – fin

TABLE HH1. Cross-classification Table for Housing and Household Weights

TABEAU LM1. Classement recoupé des coefficients de pondération des logements et des ménages

Household type — Genre de ménage	Sex of head — Sexe du chef	Age of head — Âge du chef	Number of persons — Nombre de personnes	Row No. — Rangée n°	Owned — Possédé		Rented — Loué	
					Single detached — Individuel non attenant	Other — Autres	Apartment — Appartement	Other — Autres
					Column 1 — Colonne 1	Column 2 — Colonne 2	Column 3 — Colonne 3	Column 4 — Colonne 4
One or more family households — Ménages unifamiliaux et multi-familiaux	Male — Hommes	15 - 34	LE2 – 2 ou moins	1				
			GT2 – Plus de 2	2				
		35 - 64	LE2 – 2 ou moins	3				
			GT2 – Plus de 2	4				
		65 +	LE2 – 2 ou moins	5				
			GT2 – Plus de 2	6				
	Female — Femmes	15 - 34	LE2 – 2 ou moins	7				
			GT2 – Plus de 2	8				
		35 - 64	LE2 – 2 ou moins	9				
			GT2 – Plus de 2	10				
		65 +	LE2 – 2 ou moins	11				
			GT2 – Plus de 2	12				
One person non-family households — Ménages non familiaux d'une personne	Male — Hommes	15 - 34		13				
		35 - 64		14				
		65 +		15				
	Female — Femmes	15 - 34		16				
		35 - 64		17				
		65 +		18				
Two or more person non-family households — Ménages non familiaux de deux personnes ou plus	Male — Hommes			19				
	Female — Femmes			20				

APPENDIX III

Variables and Categories Used in the Statistical Tests

A. Test of Means or Proportions

Average household size
Average age of household head
Proportion of owned households
Proportion of rented households
Proportion of dwellings that were:
 single houses;
 single houses attached to non-residential premises;

 row houses;
 duplexes;
 apartments;
 mobile.

Proportion of heads of household with mother tongue:

English;
French;
German;
Italian;
Ukrainian;
other.

B. Tests of Distributions

Variable	Category ¹
Household size	1, 2, 3, 4, 5, 6 +
Age of head	15 - 29, 30 - 44, 45 - 59, 60 +
Tenure	Owned, rented
Type of dwelling	As in (a)
Mother tongue of head	As in (b)

¹ Small categories were combined where necessary before performing the test.

APPENDICE III

Variables et catégories utilisées lors des tests statistiques

A. Test de moyennes ou de proportions

Taille moyenne des ménages
Âge moyen du chef du ménage
Proportion de logements possédés
Proportion de logements loués
Proportion de logements du type:
 logements individuels;
 logements individuels attenants à des bâtiments non résidentiels;
 maisons en rangée;
 duplex;
 appartements;
 habitations mobiles.

Proportion de chefs de ménage dont la langue maternelle est:

l'anglais;
le français;
l'allemand;
l'italien;
l'ukrainien;
autre.

B. Test de répartition

Variable	Catégorie ¹
Taille du ménage	1, 2, 3, 4, 5, 6 +
Âge du chef	15 - 29, 30 - 44, 45 - 59, 60 +
Mode d'occupation	Logement possédé, loué
Type de logement	Comme en a)
Langue maternelle du chef	Comme en b)

¹ Les petites catégories ont été agrégées au besoin avant de procéder au test.

APPENDIX – IV – APPENDICE

Differences Between Census Totals and Corresponding Sample Estimates for Selected Categories, by Province

Écarts entre les totaux du recensement et les estimations échantillonnelles correspondantes de catégories choisies, par province

Newfoundland – Terre-Neuve

Category — Catégorie	Census total — Total du recensement	Estimate from census sample — Estimation à partir de l'échantillon de recensement	Percentage difference ¹ $\frac{(2) - (1)}{(1)} \times 100\%$ Écart de pourcentage ¹ $\frac{(2) - (1)}{(1)} \times 100 \%$
	(1)	(2)	
Total:			
Households – Ménages	131,665	131,665	—
Population	557,725	557,725	—
Families – Familles	124,655	125,075	0.34
Households of size 1 - 2 – Ménages d'une ou deux personnes . .	34,050	33,995	- 0.16
Households of size 6 and over – Ménages de 6 personnes et plus	29,795	29,835	0.13
Owned households – Ménages propriétaires	106,180	106,170	- 0.01
Households occupying single dwellings – Ménages occupant des logements individuels	95,925	95,955	0.03
Household heads aged 25 - 34 – Chefs de ménage âgés de 25 - 34 ans	33,550	33,440	- 0.33
Males – Hommes	283,385	283,385	—
Females – Femmes	274,340	274,340	—
Persons aged – Personnes âgées:			
0 - 14	187,765	187,760	—
30 - 34	35,665	35,675	0.03
65 +	36,535	36,535	—
Married persons – Personnes mariées	238,670	238,665	—
Widowed persons – Veufs, veuves	20,040	20,190	0.75
Persons with English mother tongue – Personnes dont la langue maternelle est l'anglais	545,340	544,895	- 0.08
Persons with French mother tongue – Personnes dont la langue maternelle est le français	2,760	2,745	- 0.54
Two-parent families – Familles biparentales	113,853	114,270	0.36
Lone-parent families – Familles monoparentales	10,800	10,800	—
Families with no children – Familles sans enfants	26,400	26,490	0.34

See footnote(s) at end of appendix. – Voir note(s) à la fin de l'appendice.

APPENDIX IV – Continued – APPENDICE IV – suite

Differences Between Census Totals and Corresponding Sample Estimates for
Selected Categories, by Province – Continued

Écarts entre les totaux du recensement et les estimations échantillonnelles correspondantes de
catégories choisies, par province – suite

Nova Scotia – Nouvelle-Écosse

Category Catégorie	Census total Total du recensement	Estimate from census sample Estimation à partir de l'échantillon de recensement	Percentage difference ¹ $\frac{(2) - (1)}{(1)} \times 100\%$ Écart de pourcentage ¹ $\frac{(2) - (1)}{(1)} \times 100\%$
	(1)	(2)	
Total:			
Households – Ménages	243,100	243,100	–
Population	828,570	828,570	–
Families – Familles	200,480	200,640	0.08
Households of size 1 - 2 – Ménages d'une ou deux personnes . .	99,520	99,475	– 0.05
Households of size 6 and over – Ménages de 6 personnes et plus	28,425	28,070	– 1.25
Owned households – Ménages propriétaires	176,055	176,030	– 0.01
Households occupying single dwellings – Ménages occupant des logements individuels	162,550	162,470	– 0.04
Household heads aged 25 - 34 – Chefs de ménage âgés de 25 - 34 ans	53,970	53,975	–
Males – Hommes	414,155	414,155	–
Females – Femmes	414,420	414,420	–
Persons aged – Personnes âgées:			
0 - 14	223,735	223,735	–
30 - 34	53,700	53,210	– 0.91
65 +	80,735	80,730	–
Married persons – Personnes mariées	380,395	380,035	– 0.09
Widowed persons – Veufs, veuves	42,170	42,140	– 0.07
Persons with English mother tongue – Personnes dont la langue maternelle est l'anglais	768,070	767,910	– 0.02
Persons with French mother tongue – Personnes dont la langue maternelle est le français	36,870	36,880	0.03
Two-parent families – Familles biparentales	179,010	179,175	0.09
Lone-parent families – Familles monoparentales	21,470	21,465	–
Families with no children – Familles sans enfants	57,315	57,415	0.17

See footnote(s) at end of appendix. – Voir note(s) à la fin de l'appendice.

APPENDIX IV – Continued – APPENDICE IV – suite

Differences Between Census Totals and Corresponding Sample Estimates for
Selected Categories, by Province – Continued

Écarts entre les totaux du recensement et les estimations échantillonnelles correspondantes de
catégories choisies, par province – suite

Prince Edward Island – Île-du-Prince-Édouard

Category Catégorie	Census total Total du recensement	Estimate from census sample Estimation à partir de l'échantillon de recensement	Percentage difference ¹ $\frac{(2) - (1)}{(1)} \times 100\%$ Écart de pourcentage ¹ $\frac{(2) - (1)}{(1)} \times 100 \%$
	(1)	(2)	
Total:			
Households – Ménages	118,230	118,230	–
Population	32,925	32,930	–
Families – Familles	26,560	27,595	0.13
Households of size 1 - 2 – Ménages d'une ou deux personnes . .	12,770	12,740	- 0.23
Households of size 6 and over – Ménages de 6 personnes et plus	4,960	4,960	–
Owned households – Ménages propriétaires	25,225	25,225	–
Households occupying single dwellings – Ménages occupant des logements individuels	24,315	24,310	–
Household heads aged 25 - 34 – Chefs de ménage âgés de 25 - 34 ans	6,875	6,945	1.02
Males – Hommes	59,325	59,325	–
Females – Femmes	58,905	58,905	–
Persons aged – Personnes âgées:			
0 - 14	33,225	33,230	–
30 - 34	6,880	6,960	1.16
65 +	13,250	13,250	–
Married persons – Personnes mariées	52,315	52,350	0.07
Widowed persons – Veufs, veuves	6,240	6,140	- 1.60
Persons with English mother tongue – Personnes dont la langue maternelle est l'anglais	109,745	109,805	0.05
Persons with French mother tongue – Personnes dont la langue maternelle est le français	6,545	6,420	- 1.91
Two-parent families – Familles biparentales	24,685	24,720	0.14
Lone-parent families – Familles monoparentales	2,875	2,875	–
Families with no children – Familles sans enfants	7,405	7,410	–

See footnote(s) at end of appendix. – Voir note(s) à la fin de l'appendice.

APPENDIX IV – Continued – APPENDICE IV – suite

Differences Between Census Totals and Corresponding Sample Estimates for
Selected Categories, by Province – Continued

Écarts entre les totaux du recensement et les estimations échantillonnelles correspondantes de
catégories choisies, par province – suite

New Brunswick – Nouveau-Brunswick

Category Catégorie	Census total Total du recensement	Estimate from census sample Estimation à partir de l'échantillon de recensement	Percentage difference ¹ $\frac{(2) - (1)}{(1)} \times 100\%$ Écart de pourcentage ¹ $\frac{(2) - (1)}{(1)} \times 100\%$
	(1)	(2)	
Total:			
Households – Ménages	190,435	190,435	–
Population	677,250	677,245	–
Families – Familles	162,030	162,090	0.04
Households of size 1-2 – Ménages d'une ou deux personnes . .	71,875	71,585	– 0.40
Households of size 6 and over – Ménages de 6 personnes et plus	26,295	26,570	1.05
Owned households – Ménages propriétaires	136,790	136,760	– 0.02
Households occupying single dwellings – Ménages occupant des logements individuels	125,825	125,720	– 0.08
Household heads aged 25-34 – Chefs de ménage âgés de 25- 34 ans	44,210	44,420	0.48
Males – Hommes	399,335	399,335	–
Females – Femmes	337,915	337,915	–
Persons aged – Personnes âgées:			
0-14	193,100	193,095	–
30-34	42,990	42,680	– 0.72
65+	61,075	61,075	–
Married persons – Personnes mariées	309,345	309,085	– 0.08
Widowed persons – Veufs, veuves	30,750	31,125	1.22
Persons with English mother tongue – Personnes dont la langue maternelle est l'anglais	435,975	435,480	– 0.07
Persons with French mother tongue – Personnes dont la langue maternelle est le français	223,780	223,715	– 0.03
Two-parent families – Familles biparentales	145,876	145,950	0.05
Lone-parent families – Familles monoparentales	16,160	16,145	– 0.09
Families with no children – Familles sans enfants	43,425	43,470	0.10

See footnote(s) at end of appendix. – Voir note(s) à la fin de l'appendice.

APPENDIX IV – Continued – APPENDICE IV – suite

Differences Between Census Totals and Corresponding Sample Estimates for
Selected Categories, by Province – Continued

Écarts entre les totaux du recensement et les estimations échantillonnelles correspondantes de
catégories choisies, par province – suite

Québec

Category Catégorie	Census total Total du recensement	Estimate from census sample Estimation à partir de l'échantillon de recensement	Percentage difference ¹ $\frac{(2) - (1)}{(1)} \times 100\%$ Écart de pourcentage ¹ $\frac{(2) - (1)}{(1)} \times 100\%$
	(1)	(2)	(1)
Total:			
Households – Ménages	1,894,110	1,894,115	–
Population	6,234,445	6,234,450	–
Families – Familles	1,540,400	1,542,645	0.15
Households of size 1 - 2 – Ménages d'une ou deux personnes . .	796,220	796,315	0.01
Households of size 6 and over – Ménages de 6 personnes et plus	191,095	190,050	– 0.55
Owned households – Ménages propriétaires	953,955	953,840	– 0.01
Households occupying single dwellings – Ménages occupant des logements individuels	745,595	745,380	– 0.03
Household heads aged 25 - 34 – Chefs de ménage âgés de 25 - 34 ans	472,890	473,025	0.03
Males – Hommes	3,084,650	3,084,650	–
Females – Femmes	3,149,800	3,149,800	–
Persons aged – Personnes âgées:			
0 - 14	1,550,335	1,550,340	–
30 - 34	474,555	473,775	– 0.16
65 +	481,360	481,355	–
Married persons – Personnes mariées	2,906,970	2,905,085	– 0.06
Widowed persons – Veufs, veuves	259,945	261,115	0.45
Persons with English mother tongue – Personnes dont la langue maternelle est l'anglais	800,680	797,420	– 0.40
Persons with French mother tongue – Personnes dont la langue maternelle est le français	4,989,245	4,988,545	– 0.01
Two-parent families – Familles biparentales	1,381,506	1,383,870	0.17
Lone-parent families – Familles monoparentales	158,900	158,775	– 0.08
Families with no children – Familles sans enfants	427,845	428,565	0.17

See footnote(s) at end of appendix. – Voir note(s) à la fin de l'appendice.

APPENDIX IV – Continued – APPENDICE IV – suite

Differences Between Census Totals and Corresponding Sample Estimates for
Selected Categories, by Province – Continued

Écarts entre les totaux du recensement et les estimations échantillonnelles correspondantes de
catégories choisies, par province – suite
Ontario

Category Catégorie	Census total Total du recensement	Estimate from census sample Estimation à partir de l'échantillon de recensement	Percentage difference ¹ $\frac{(2) - (1)}{(1)} \times 100\%$ Écart de pourcentage ¹ $\frac{(2) - (1)}{(1)} \times 100\%$
	(1)	(2)	
Total:			
Households – Ménages	2,634,620	2,634,620	–
Population	8,264,465	8,264,465	–
Families – Familles	2,104,545	2,106,295	0.08
Households of size 1 - 2 – Ménages d'une ou deux personnes . .	1,193,630	1,194,350	0.06
Households of size 6 and over – Ménages de 6 personnes et plus	212,615	211,020	– 0.75
Owned households – Ménages propriétaires	1,676,250	1,676,265	–
Households occupying single dwellings – Ménages occupant des logements individuels	1,494,465	1,494,015	– 0.03
Household heads aged 25 - 34 – Chefs de ménage âgés de 25 - 34 ans	607,720	608,265	0.09
Males – Hommes	4,096,865	4,096,865	–
Females – Femmes	4,167,600	4,167,600	–
Persons aged – Personnes âgées:			
0 - 14	2,073,785	2,073,780	–
30 - 34	586,445	584,270	– 0.37
65 +	738,915	738,920	–
Married persons – Personnes mariées	4,052,510	4,050,480	– 0.05
Widowed persons – Veufs, veuves	395,495	397,035	0.39
Persons with English mother tongue – Personnes dont la langue maternelle est l'anglais	6,457,645	6,456,345	– 0.02
Persons with French mother tongue – Personnes dont la langue maternelle est le français	462,070	461,625	– 0.10
Two-parent families – Familles biparentales	1,902,090	1,903,950	0.10
Lone-parent families – Familles monoparentales	202,450	202,350	– 0.05
Families with no children – Familles sans enfants	650,990	651,700	0.11

See footnote(s) at end of appendix. – Voir note(s) à la fin de l'appendice.

APPENDIX IV – Continued – APPENDICE IV – suite

Differences Between Census Totals and Corresponding Sample Estimates for
Selected Categories, by Province – Continued

Écarts entre les totaux du recensement et les estimations échantillonnelles correspondantes de
catégories choisies, par province – suite

Manitoba

Category — Catégorie	Census total — Total du recensement	Estimate from census sample — Estimation à partir de l'échantillon de recensement	Percentage difference ¹ $\frac{(2) - (1)}{(1)} \times 100\%$ — Écart de pourcentage ¹ $\frac{(2) - (1)}{(1)} \times 100\%$
	(1)	(2)	
Total:			
Households – Ménages	328,005	328,005	—
Population	1,021,505	1,021,510	—
Families – Familles	251,975	252,305	0.13
Households of size 1-2 – Ménages d'une ou deux personnes . .	159,290	159,530	0.15
Households of size 6 and over – Ménages de 6 personnes et plus	27,575	26,900	– 2.45
Owned households – Ménages propriétaires	217,685	217,665	– 0.01
Households occupying single dwellings – Ménages occupant des logements individuels	219,950	219,960	—
Household heads aged 25-34 – Chefs de ménage âgés de 25-34 ans	70,000	70,155	0.22
Males – Hommes	508,010	508,010	—
Females – Femmes	513,495	513,495	—
Persons aged – Personnes âgées:			
0-14	265,105	265,105	—
30-34	65,450	65,195	– 0.39
65+	106,560	106,565	—
Married persons – Personnes mariées	484,815	484,235	– 0.12
Widowed persons – Veufs, veuves	52,615	53,020	0.77
Persons with English mother tongue – Personnes dont la langue maternelle est l'anglais	727,240	727,065	– 0.02
Persons with French mother tongue – Personnes dont la langue maternelle est le français	54,745	54,275	– 0.86
Two-parent families – Familles biparentales	227,241	227,560	0.14
Lone-parent families – Familles monoparentales	24,735	24,750	0.06
Families with no children – Familles sans enfants	81,305	81,490	0.23

See footnote(s) at end of appendix. – Voir note(s) à la fin de l'appendice.

APPENDIX IV – Continued – APPENDICE IV – suite

Differences Between Census Totals and Corresponding Sample Estimates for
Selected Categories, by Province – Continued

Écarts entre les totaux du recensement et les estimations échantillonnelles correspondantes de
catégories choisies, par province – suite

Saskatchewan

Category – Catégorie	Census total – Total du recensement (1)	Estimate from census sample – Estimation à partir de l'échantillon de recensement (2)	Percentage difference ¹ $\frac{(2) - (1)}{(1)} \times 100\%$ Écart de pourcentage ¹ $\frac{(2) - (1)}{(1)} \times 100 \%$
Total:			
Households – Ménages	291,150	291,150	–
Population	921,325	921,320	–
Families – Familles	225,685	225,960	0.12
Households of size 1 - 2 – Ménages d'une ou deux personnes . .	139,860	139,925	0.05
Households of size 6 and over – Ménages de 6 personnes et plus	28,045	28,120	0.26
Owned households – Ménages propriétaires	219,925	219,945	0.01
Households occupying single dwellings – Ménages occupant des logements individuels	224,510	224,290	– 0.10
Household heads aged 25 - 34 – Chefs de ménage âgés de 25 - 34 ans	52,840	53,230	0.74
Males – Hommes	464,775	464,775	–
Females – Femmes	456,550	456,550	–
Persons aged – Personnes âgées:			
0 - 14	248,015	248,015	–
30 - 34	50,960	50,795	– 0.32
65 +	102,170	102,160	– 0.01
Married persons – Personnes mariées	435,870	435,695	– 0.04
Widowed persons – Veufs, veuves	47,445	47,930	1.02
Persons with English mother tongue – Personnes dont la langue maternelle est l'anglais	715,685	715,270	– 0.06
Persons with French mother tongue – Personnes dont la langue maternelle est le français	26,710	26,540	– 0.64
Two-parent families – Familles biparentales	206,584	206,860	0.13
Lone-parent families – Familles monoparentales	19,105	19,100	–
Families with no children – Familles sans enfants	73,745	74,025	0.38

See footnote(s) at end of appendix. – Voir note(s) à la fin de l'appendice.

APPENDIX IV - Continued - APPENDICE IV - suite

Differences Between Census Totals and Corresponding Sample Estimates for
Selected Categories, by Province - Continued

Écarts entre les totaux du recensement et les estimations échantillonnelles correspondantes de
catégories choisies, par province - suite

Alberta

Category Catégorie	Census total Total du recensement (1)	Estimate from census sample Estimation à partir de l'échantillon de recensement (2)	Percentage difference ¹ $\frac{(2) - (1)}{(1)} \times 100\%$ Écart de pourcentage ¹ $\frac{(2) - (1)}{(1)} \times 100 \%$
Total:			
Households - Ménages	575,280	575,280	-
Population	1,838,035	1,838,035	-
Families - Familles	448,770	449,010	0.05
Households of size 1-2 - Ménages d'une ou deux personnes . .	260,125	260,585	0.18
Households of size 6 and over - Ménages de 6 personnes et plus	49,165	49,385	0.45
Owned households - Ménages propriétaires	372,825	372,820	-
Households occupying single dwellings - Ménages occupant des logements individuels	372,420	372,120	- 0.08
Household heads aged 25-34 - Chefs de ménage âgés de 25-34 ans	141,585	142,370	0.55
Males - Hommes	932,370	932,370	-
Females - Femmes	905,670	905,670	-
Persons aged - Personnes âgées:			
0-14	503,130	503,125	-
30-34	128,815	128,345	- 0.36
65+.	137,925	137,925	-
Married persons - Personnes mariées	865,465	864,945	- 0.06
Widowed persons - Veufs, veuves	69,410	69,805	0.57
Persons with English mother tongue - Personnes dont la langue maternelle est l'anglais	1,482,725	1,482,015	- 0.05
Persons with French mother tongue - Personnes dont la langue maternelle est le français	44,440	44,555	0.26
Two-parent families - Familles biparentales	407,570	407,835	0.07
Lone-parent families - Familles monoparentales	41,200	41,170	- 0.07
Families with no children - Familles sans enfants	135,110	135,345	0.17

See footnote(s) at end of appendix. - Voir note(s) à la fin de l'appendice.

APPENDIX IV – Concluded – APPENDICE IV – fin

Differences Between Census Totals and Corresponding Sample Estimates for
Selected Categories, by Province – Concluded

Écarts entre les totaux du recensement et les estimations échantillonnelles correspondantes de
catégories choisies, par province – fin

British Columbia – Colombie-Britannique

Category — Catégorie	Census total — Total du recensement	Estimate from census sample — Estimation à partir de l'échantillon de recensement	Percentage difference ¹ $\frac{(2) - (1)}{(1)} \times 100\%$ Écart de pourcentage ¹ $\frac{(2) - (1)}{(1)} \times 100\%$
	(1)	(2)	
Total:			
Households – Ménages	828,285	828,285	–
Population	2,466,605	2,466,610	–
Families – Familles	628,445	629,345	0.14
Households of size 1-2 – Ménages d'une ou deux personnes. .	422,125	422,675	0.13
Households of size 6 and over – Ménages de 6 personnes et plus	55,130	54,450	– 1.23
Owned households – Ménages propriétaires	540,635	540,590	– 0.01
Households occupying single dwellings – Ménages occupant des logements individuels	516,485	516,255	– 0.04
Household heads aged 25-34 – Chefs de ménage âgés de 25-34 ans	189,770	189,430	– 0.18
Males – Hommes	1,232,510	1,232,510	–
Females – Femmes	1,234,095	1,234,095	–
Persons aged – Personnes âgées:			
0-14	595,125	595,125	–
30-34	176,875	176,335	– 0.31
65+	242,055	242,050	–
Married persons – Personnes mariées	1,221,535	1,221,190	– 0.03
Widowed persons – Veufs, veuves	118,230	118,385	0.13
Persons with English mother tongue – Personnes dont la langue maternelle est l'anglais	2,037,645	2,037,020	– 0.03
Persons with French mother tongue – Personnes dont la langue maternelle est le français	38,430	37,895	– 1.39
Two-parent families – Familles biparentales	568,248	569,160	0.16
Lone-parent families – Familles monoparentales.	60,200	60,185	– 0.02
Families with no children – Familles sans enfants	219,985	220,425	0.20

¹ Differences of ± 5 are neglected since they could arise from random rounding. – Il n'est pas tenu compte des écarts de ± 5 étant donné qu'ils peuvent découler d'un arrondissement aléatoire.

REFERENCES – RÉFÉRENCES

- Brackstone, G.J. and Hill, C.J. The Estimation of Total Variance in the 1976 Census, *Survey Methodology Journal*, Vol. 2, No. 2, Statistics Canada, 1976.
- . *The Estimation of Total Variance in the 1976 Census*, *Survey Methodology Journal*, vol. 2, n° 2, Statistique Canada, 1976.
- Canada. Dominion Bureau of Statistics. *Eighth Census of Canada, 1941, Administrative Report of the Dominion Statistician*. Ottawa: King's Printer, 1945.
- . Bureau fédéral de la statistique. *Huitième recensement du Canada, 1941, Rapport administratif du Statisticien fédéral*. Ottawa: Imprimeur du Roi, 1945.
- . Dominion Bureau of Statistics. *Ninth Census of Canada, 1951, Vol. XI, Administrative Report*. Ottawa: Queen's Printer, 1955.
- . Bureau fédéral de la statistique. *Neuvième recensement du Canada, 1951, vol. XI, Rapport administratif*. Ottawa: Imprimeur de la Reine, 1955.
- . Dominion Bureau of Statistics. *1961 Census of Canada, General Review*, Bulletin 7.2-12, Catalogue 99-537. Ottawa: Queen's Printer, 1970.
- . Bureau fédéral de la statistique. *Recensement du Canada de 1961, Revue générale*, bulletin 7.2-12, n° 99-537 au catalogue. Ottawa: Imprimeur de la Reine, 1970.
- . Dominion Bureau of Statistics. *Sampling in the Census, S.M.S.03.5*, August 1968.
- . Bureau fédéral de la statistique. *Sampling in the Census S.M.S.03.5*, août 1968.
- . Statistics Canada. *1971 Census of Canada, General Review, Vol. VI, Part 1*, Catalogue 99-740. Ottawa, 1976.
- . Statistique Canada. *Recensement du Canada de 1971, Revue générale, vol. VI, partie 1*, n° 99-740 au catalogue. Ottawa, 1976.
- Cochran, W.G. *Sampling Techniques (Third Edition)*. New York, John Wiley & Sons, 1977.
- Cochran, W.G. *Sampling Techniques (troisième édition)*. New York, John Wiley & Sons, 1977.
- Fellegi, I.P. Response Variance and its Estimation, *Journal of the American Statistical Association*, 59 (December 1964), pp. 1016-1041.
- . *Response Variance and its Estimation*, *Journal of the American Statistical Association*, n° 59 (décembre 1964), pp. 1016-1041.
- Hansen, M.H., Hurwitz, W.N. and Bershad, M.A. Measurement Errors in Censuses and Surveys, *Bulletin of the International Statistical Institute*, 38 (August 1959), pp. 359-374.
- . *Measurement Errors in Censuses and Surveys*, *Bulletin of the International Statistical Institute*, n° 38 (août 1959), pp. 359-374.
- Stuart, A. *Basic Ideas of Scientific Sampling (Second Edition)*, New York: Hafner Press, 1976.
- . *Basic Ideas of Scientific Sampling (deuxième édition)*. New York: Hafner Press, 1976.
- U.S. Bureau of the Census. *Technical Paper No. 13, Sampling Applications in Censuses of Population and Housing*. Washington, D.C., 1965.
- . *Technical Paper No. 13, Sampling Applications in Censuses of Population and Housing*. Washington, D.C., 1965.

Ca OOS

Statistics Canada Library
Bibliothèque Statistique Canada

1010021204

DATE DUE
DATE DE RETOUR

MAY MAI	16	1980		
AUG AOUT	21	1980		
FEB FEV	12	1981		
MAY MAI	4	1981		
		<i>V. Henneville</i>		
JUN JUIN	10	1981		
FEB FEV	11	1982	<i>Coyne</i>	
APR AVR	28	1982	<i>Brange</i>	
.....			<i>Coyne</i>	
AUG AOUT	19	1983	<i>Royce</i>	
AUG AOUT	10	1987		
		<i>Raithes</i>		
FEB FEV	-	1992	<i>X</i>	
LOWE-MARTIN No. 1137				

